Audiovisual Processing of Chinese Characters Elicits Suppression and Congruency Effects in MEG

Weiyong Xu [1,2]*, Orsolya Beatrix Kolozsvári [1,2], Robert Oostenveld [3,4], Paavo Herman Tapio Leppänen [1,2] and Jarmo Arvid Hämäläinen [1,2]

[1]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, [2]Jyväskylä Centre for Interdisciplinary Brain Research, Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, [4]NatMEG, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Learning to associate written letters/characters with speech sounds is crucial for reading acquisition. Most previous studies have focused on audiovisual integration in alphabetic languages. Less is known about logographic languages such as Chinese characters, which map onto mostly syllable-based morphemes in the spoken language. Here we investigated how long-term exposure to native language affects the underlying neural mechanisms of audiovisual integration in a logographic language using magnetoencephalography (MEG). MEG sensor and source data from 12 adult native Chinese speakers and a control group of 13 adult Finnish speakers were analyzed for audiovisual suppression (bimodal responses vs. sum of unimodal responses) and congruency (bimodal incongruent responses vs. bimodal congruent responses) effects. The suppressive integration effect was found in the left angular and supramarginal gyri (205–365 ms), left inferior frontal and left temporal cortices (575–800 ms) in the Chinese group. The Finnish group showed a distinct suppression effect only in the right parietal and occipital cortices at a relatively early time window (285–460 ms). The congruency effect was only observed in the Chinese group in left inferior frontal and superior temporal cortex in a late time window (about 500–800 ms) probably related to modulatory feedback from multi-sensory regions and semantic processing. The audiovisual integration in a logographic language showed a clear resemblance to that in alphabetic languages in the left superior temporal cortex, but with activation specific to the logographic stimuli observed in the left inferior frontal cortex. The current MEG study indicated that learning of logographic languages has a large impact on the audiovisual integration of written characters with some distinct features compared to previous results on alphabetic languages.

Keywords: audiovisual integration, magnetoencephalography, auditory cortex, language learning, reading, Chinese characters

INTRODUCTION

Learning to read involves the integration of multisensory information (primarily from the auditory and visual modalities) and combining it with meaning. Multisensory integration, defined as modulation of brain responses by signals from multiple modalities, has been shown to be a dynamic and context-dependent process (van Atteveldt et al., 2014; Murray et al., 2016a). In natural audiovisual speech perception, it has been shown that complementary or correlated visual speech information could affect the auditory processing of speech by activating the perisylvian auditory speech regions through the ventral and dorsal visual streams (Campanella and Belin, 2007; Campbell, 2008; Bernstein and Liebenthal, 2014). Multisensory integration is also found within the visual cortices during early and late post-stimulus stages and could directly impact perception (Murray et al., 2016b; Kayser et al., 2017). Unlike spoken language, the ability to read is not hard-wired in the human brain through the evolution since written language is a recent cultural invention which has only existed for a few thousand years (Liberman, 1992). Consequently, it takes years of repetition and practice to form the long-term memory representations of audiovisual language objects, which would enable fluent readers to successfully automatize the integration of language-related auditory and visual sensory information (Froyen et al., 2009). A growing body of neuroimaging research has examined the neurophysiological basis of the letter-speech sound integration mainly in transparent alphabetic languages (Raij et al., 2000; van Atteveldt et al., 2004, 2009; Froyen et al., 2009). Interestingly, research using less transparent language such as English has found both similarities in the audiovisual integration effect compared with transparent languages and orthography dependent differences in the processing of irregular mappings of letter-speech sound combinations (Holloway et al., 2015). Therefore, an intriguing open question is about the audiovisual integration in other kinds of languages, for example character-speech processing in logographic languages such as Chinese. Understanding of such character-speech integration in logographic languages may provide more insights into the universal and language-specific brain circuits underlying audiovisual integration in reading acquisition.

Audiovisual paradigms, which typically consisted of auditory only (A), visual only (V), audiovisual congruent (AVC) and audiovisual incongruent (AVI) stimuli, were widely used in investigating brain mechanisms of multisensory interactions between the auditory and visual modalities (Raij et al., 2000; Murray and Spierer, 2009; van Atteveldt et al., 2009). Within the structure of such experimental design, two main approaches could be derived and used as indications of audiovisual integration. The first approach is based on the additive model, which compares the audiovisual responses to the summations of the constituent unisensory responses [AV − (A + V)], and has been frequently used in electrophysiological studies on multisensory integration (Raij et al., 2000; Calvert and Thesen, 2004; Stein and Stanford, 2008; Sperdin et al., 2009). The additive model could be applied to almost any kind of

audiovisual experimental design with arbitrary auditory and visual combinations. This approach is suitable to detect both supra-additive [AV > (A + V)] and sub-additive (hereon referred to as the suppression effect) [AV < (A + V)] modulations of unimodal activities in the sensory-specific cortices as well as to observe new processes specifically activated by the bimodal nature of the stimulus under the assumption that there is no common activity (such as target processing) presented in the auditory, visual and audiovisual conditions (Besle et al., 2004). Animal electrophysiological studies have shown both super- and sub-additive multisensory interactions in the superior colliculus neurons and the superior temporal sulcus (STS; Meredith, 2002; Schroeder and Foxe, 2002; Laurienti et al., 2005; Perrault et al., 2005; Kayser et al., 2008; Stein and Stanford, 2008). Electroencephalography (EEG)/magnetoencephalography (MEG) studies on humans have typically shown suppressive multisensory effects (Schröger and Widmann, 1998; Foxe et al., 2000; Raij et al., 2000; Fort et al., 2002; Lütkenhöner et al., 2002; Molholm et al., 2002; Teder-Sälejärvi et al., 2002; Jost et al., 2014; Xu et al., 2018). Such suppression effects could occur as early as 50–60 ms after the stimulus onset and these functionally coupled responses are localized within the primary visual and auditory cortices as well as the posterior STS (Cappe et al., 2010). Other research in humans has also demonstrated that these audiovisual suppression effects can be observed at late time windows and for both familiar and unfamiliar audiovisual stimuli (Raij et al., 2000; Jost et al., 2014; Xu et al., 2018).

The second approach is to study the audiovisual congruency effect (Jones and Callan, 2003; Ojanen et al., 2005; Hein et al., 2007; Rüsseler et al., 2018), which compares different brain responses to congruent and incongruent audiovisual pairs. The rationale is that the congruency effect can only be established when the unisensory inputs have been integrated successfully (van Atteveldt et al., 2007a,b). One advantage of the congruency comparison is that it is not sensitive to any additional non-sensory activity and thus has a clear statistical criterion. Research using transparent languages such as Dutch and Finnish has found that the congruent alphabetic audiovisual stimuli elicit a stronger brain response in the superior temporal cortex than the incongruent pairs (Raij et al., 2000; van Atteveldt et al., 2004). Based on earlier studies (Raij et al., 2000; Besle et al., 2004; Cappe et al., 2010; Jost et al., 2014), the suppression effect reflects general audiovisual integration (including both early and late audiovisual interaction effects), whereas the congruency effect is more related to the specific interaction of learned or meaningful audiovisual associations (Hocking and Price, 2009).

Previous research has mainly focused on the brain mechanisms of audiovisual integration in alphabetic languages such as Dutch (van Atteveldt et al., 2004, 2009), English (Holloway et al., 2015) and Finnish (Raij et al., 2000). Several language-related and cross-modal brain regions have been shown to activate consistently during letter-speech sound integration in alphabetic languages. In particular, the superior temporal cortices have been reported in fMRI studies to have heteromodal properties (van Atteveldt et al., 2004, 2009; Blau et al., 2008). The left and right STS have also been implicated to be the main letter-speech sound integration regions in

an early MEG study using Finnish letters (Raij et al., 2000). In addition, feedback projections from these cross-modal regions were found to alter the brain activities in the primary auditory cortex (van Atteveldt et al., 2004). The heteromodal areas in the temporal cortices have shown differences in their tolerances to temporal synchrony between modalities: the visual and auditory inputs are integrated in the STS and superior temporal gyrus (STG) within a broad range of temporal cross-modal synchrony between the auditory and visual stimuli, while the effect of congruent and incongruent auditory-visual stimuli rapidly diminishes with decreasing temporal synchrony in planum temporale (PT) and Heschl's sulcus (HS) regions (van Atteveldt et al., 2007a). Top-down influences by different instructions and task demands also evidently affect the congruency effects (Andersen et al., 2004). For instance, different experimental designs (explicit/implicit and active/passive) have been shown to modulate the letter-speech sound congruency effect in fMRI (van Atteveldt et al., 2007b; Blau et al., 2008).

While fMRI provides accurate locations of the integration sites, its poor temporal resolution fails to capture the timing information. EEG and MEG have an excellent temporal resolution (in millisecond scale) and could provide additional timing information about audiovisual integration processes. An early MEG study using Finnish letters and speech sounds revealed that the auditory and visual sensory inputs showed maximal activities in multimodal sites about 225 ms after stimulus onset (Raij et al., 2000). It was followed by suppressive interaction at 280–345 ms in the right temporo-occipito-parietal junction and at 380–540 ms in the left STS and 450–535 ms in the right STS (Raij et al., 2000). Another MEG study using Hiragana graphemes and phonemes found that congruent audiovisual stimuli evoked larger 2–10 Hz oscillations in the left auditory cortex within the first 250 ms after stimulus onset and smaller 2–16 Hz oscillations in bilateral visual cortices between 250 and 500 ms compared with incongruent audiovisual stimuli (Herdman et al., 2006). Corroborating evidence comes from an EEG study that found audiovisual suppression effects in event-related potentials (ERPs) for familiar German (300–324 and 480–764 ms) and unfamiliar English words (324–384 and 416–756 ms), whereas audiovisual congruency effect could be found only for familiar German words (160–204, 544–576, 1032–1108 and 1164–1188 ms), and it was characterized by topographic differences probably due to lexical-semantic processes involved (Jost et al., 2014).

However, most studies on audiovisual integration have only investigated letter-speech sound mappings in alphabetic languages, and less is known about the logographic language such as Chinese. In Chinese, a single character encodes morphosyllabic information that represents a syllable with a distinct meaning (Perfetti and Tan, 1998; Shu, 2003; Tan et al., 2005; Ziegler and Goswami, 2005; McBride, 2015). For native Chinese speakers, long-term memory representations of characters through years of learning and reading repetition would enable them to retrieve the phonological and semantic information embedded in the characters directly. Semantic processing has been shown to modulate the brain activity around 400 ms, with the N400 response being sensitive to semantic incongruency (Kutas and Federmeier, 2011; Du et al., 2014; Jost et al., 2014). The magnetic counterpart of the N400 (N400m) has also been shown to be sensitive to semantic content and to have sources in the bilateral temporal cortices in unimodal conditions (Service et al., 2007; Vartiainen et al., 2009). Thus, research using logographic language such as Chinese is important to advance our understanding of the general mechanisms of audiovisual integration including cross-modal semantic congruency processing in the human brain.

In the current study, we investigated the dynamics of cortical activation to logographic multisensory stimuli using MEG. We designed an active one-back audiovisual cover task during which audiovisual integration could be examined. We presented Chinese characters and speech sounds as stimuli in A, V, AVC, and AVI conditions to native speakers of Chinese and used native speakers of Finnish, who were naive to Chinese, as a control group to verify the effects of long-term exposure to these stimuli. The MEG data were analyzed at both sensor and source levels to examine the suppression and congruency effects. Highly automatized audiovisual associations could affect processes related to general cross-modal integration manifested as suppression effect and congruency detection. More specifically, we hypothesized that learned characters combined with the corresponding congruent/incongruent speech sounds would elicit an audiovisual congruency effect manifested as a combined modulatory feedback and semantic N400m response, but only in native Chinese speakers. In addition, we examined the effects of long-term exposure of logographic language on the audiovisual suppressive interaction (sum of auditory and visual only conditions compared to audiovisual condition) which should be affected less by semantic processing than the congruency effect. Instead, suppressive integration is likely to reflect more general cross-modal processes. We thus expected more similar suppression than congruency effect to alphabetic languages found in earlier studies.

MATERIALS AND METHODS

Participants

Chinese participants were adults and native speakers of Mandarin Chinese studying in Jyväskylä, Finland. Chinese participants had learned simplified Chinese characters through formal education. In addition, another group of native speakers of Finnish was recruited as a control group. Finnish participants had no prior exposure to the Chinese language. All participants included had normal hearing and normal or corrected-to-normal vision. They were screened for the following exclusion criteria: head injuries, ADHD, neurological diseases, medication affecting the central nervous system, language delays or any other language-related disorders. Ethical approval for this study was obtained from the Ethics Committee of the University of Jyväskylä. This study was carried out in accordance with the recommendations of the Ethics Committee of the University of Jyväskylä with written informed consent from

all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee of the University of Jyväskylä. After the MEG experiments, all of them received movie tickets as compensations for their time. In total, we measured MEG data from 19 native Chinese speakers and 18 Finnish speakers. Of those seven were excluded from the Chinese group and five were excluded in the Finnish group for following reasons: four subjects due to poor visual acuity even with magnetically neutral glasses for vision correction, two subjects due to excessive head movements or low head positions in the MEG helmet, one due to technical problems during the recording, and five subjects due to strong noise interference and poor signal quality. The final dataset thus included 12 Chinese participants and 13 Finnish participants (**Table 1**).

Stimuli and Task

The stimuli consisted of six Chinese characters (Simplified Chinese) and their corresponding flat tone speech sounds (1. 步: bu; 2. 都: du; 3. 谷: gu; 4. 酷: ku; 5. 普: pu; 6. 兔: tu). The characters were all common characters familiar to the native Chinese speakers and had the following meanings: 1. steps/walking; 2. both/capital; 3. grain; 4. cool; 5. common; and 6. rabbit. The mean duration of the auditory stimuli was 447.2 ms (SD: 32.7 ms). The duration of the visual stimuli was 1,000 ms. Four kinds of stimuli, A, V, AVC and AVI were presented in random order with 108 trials in each type of the stimuli. Each trial started with a fixation cross at the center of the screen for 1,000 ms. In the AVI and AVC conditions, there was a 36 ms delay between the visual and auditory stimuli. This time delay is within the optimal range of cross-modal integration (Schroeder and Foxe, 2002; Kayser et al., 2008). The visual stimuli were projected onto the center of the screen one meter away from the participants with a white color background. The fixation cross was 1.2 cm and the characters were 2.5 × 2.5 cm on the screen. The stimuli were presented with Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA) software running in a Microsoft Windows computer.

Participants were instructed to do a dual-modality one-back task in order to keep their attention equally on both auditory and visual stimuli (**Figure 1**). Cover task trials as shown in **Figure 1** occurred randomly (with 7.5% probability across all trials) to keep their attention on both auditory and visual modalities. The cover task trials tested the participant's memory about the auditory and visual stimuli one trial before the last trial. The cover task trials consisted of one test trial for the auditory stimulus, one test trial for the visual stimulus followed by feedback. Four options were given on the screen in each
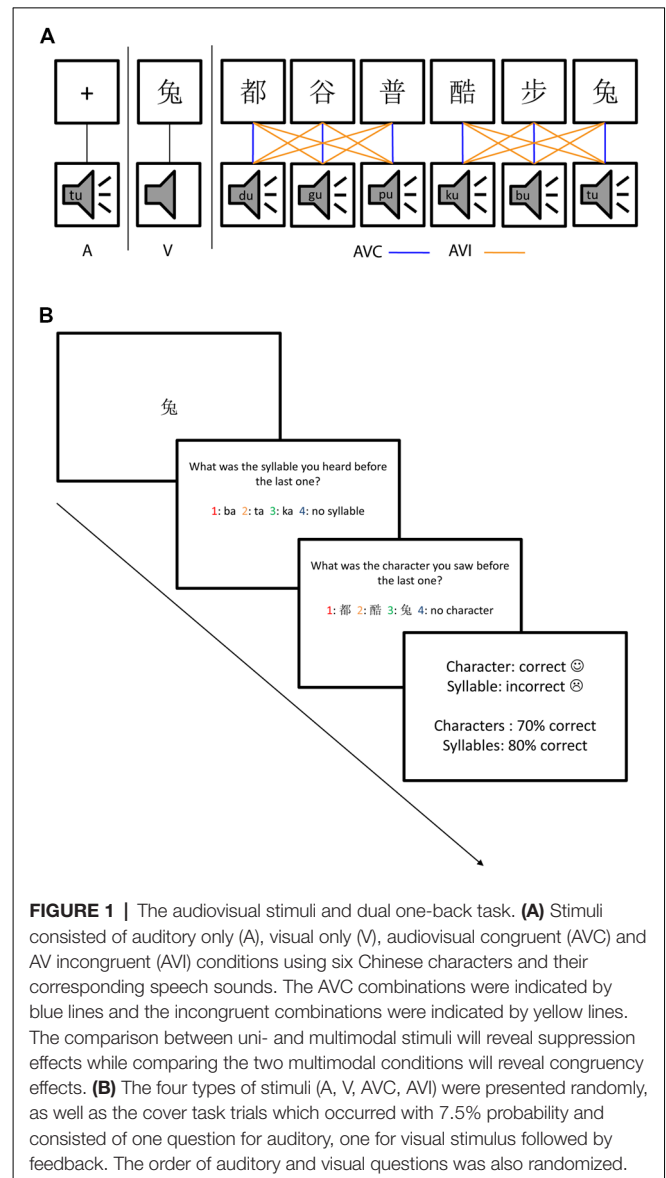

FIGURE 1 | The audiovisual stimuli and dual one-back task. **(A)** Stimuli consisted of auditory only (A), visual only (V), audiovisual congruent (AVC) and AV incongruent (AVI) conditions using six Chinese characters and their corresponding speech sounds. The AVC combinations were indicated by blue lines and the incongruent combinations were indicated by yellow lines. The comparison between uni- and multimodal stimuli will reveal suppression effects while comparing the two multimodal conditions will reveal congruency effects. **(B)** The four types of stimuli (A, V, AVC, AVI) were presented randomly, as well as the cover task trials which occurred with 7.5% probability and consisted of one question for auditory, one for visual stimulus followed by feedback. The order of auditory and visual questions was also randomized.

of the auditory or visual test trials and the participants needed to choose the correct one from the four options using the response pad. The order of the auditory and visual test trials was randomized.

MEG Recording

MEG data (102 magnetometer channels and 204 planar gradiometer channels; sampling rate: 1,000 Hz; band-pass filter: 0.1–330 Hz) were recorded using Elekta Neuromag® TRIUX™ system (Elekta AB, Stockholm, Sweden) in a magnetically shielded room. The head position with respect to the MEG sensors in the helmet was monitored continuously with five digitized head position indicator (HPI) coils attached to the scalp. Three HPI coils were placed on the forehead and one behind each ear. At the beginning of each MEG recording, the positions of HPI coils were determined by three anatomic landmarks (nasion, left and right preauricular points), which were digitized using the

TABLE 1 | Demographic information of the study participants.

Group	Chinese	Finnish
Number of participants	12	13
Age	24.36 ± 3.66	24.31 ± 2.06
Sex	9 Female, 3 Male	9 Female, 4 Male
Handedness	12 right	12 right, 1 ambidextrous

Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, USA). To allow the co-registration with the MRI template, an additional set of points (>100) randomly distributed over the scalp were also digitized. Electro-oculogram (EOG) signal was recorded with one ground electrode attached to the collarbone and another two diagonally placed electrodes (one slightly above the right eye and one slightly below the left eye).

Data Analysis

First, MEG data were processed with Maxfilter 3.0^{TM} (Elekta AB, Stockholm, Sweden) to remove external magnetic disturbance and correct for head movements (Taulu et al., 2004; Taulu and Kajola, 2005; Taulu and Simola, 2006). Bad channels were marked manually and were excluded and later reconstructed in the Maxfilter. The temporal extension of the signal-space separation method (tSSS) was applied in buffers of 30 s (Taulu et al., 2004; Taulu and Kajola, 2005; Taulu and Simola, 2006). The head position was estimated using 200 ms time windows with 10 ms steps for head movement compensation. The mean head position across the whole MEG recording session was used for head position transformation.

MEG data were then analyzed using MNE Python (0.14; Gramfort et al., 2013). First, a low-pass filter of 40 Hz (firwin2 filter design, transition bandwidth 10 Hz) was applied to the continuous MEG recordings and data were segmented into epochs −200 to 1,000 ms relative to the stimulus onset. Data were then manually checked to excluded any trials that contaminated by head movement related artifacts or electronic jump artifacts. Then fast independent component analysis (ICA) was applied to remove any eye blink and cardiac artifacts related components. MEG epochs exceeding 1 pT/cm (for gradiometer channels) or 3pT (for magnetometer channels) peak-to-peak amplitudes were excluded from further analysis. Event-related fields (ERFs) were obtained by averaging trials in the four conditions (A, V, AVC, and AVI) separately. Sum of auditory and visual responses (A + V) was calculated by adding up the auditory and visual ERFs
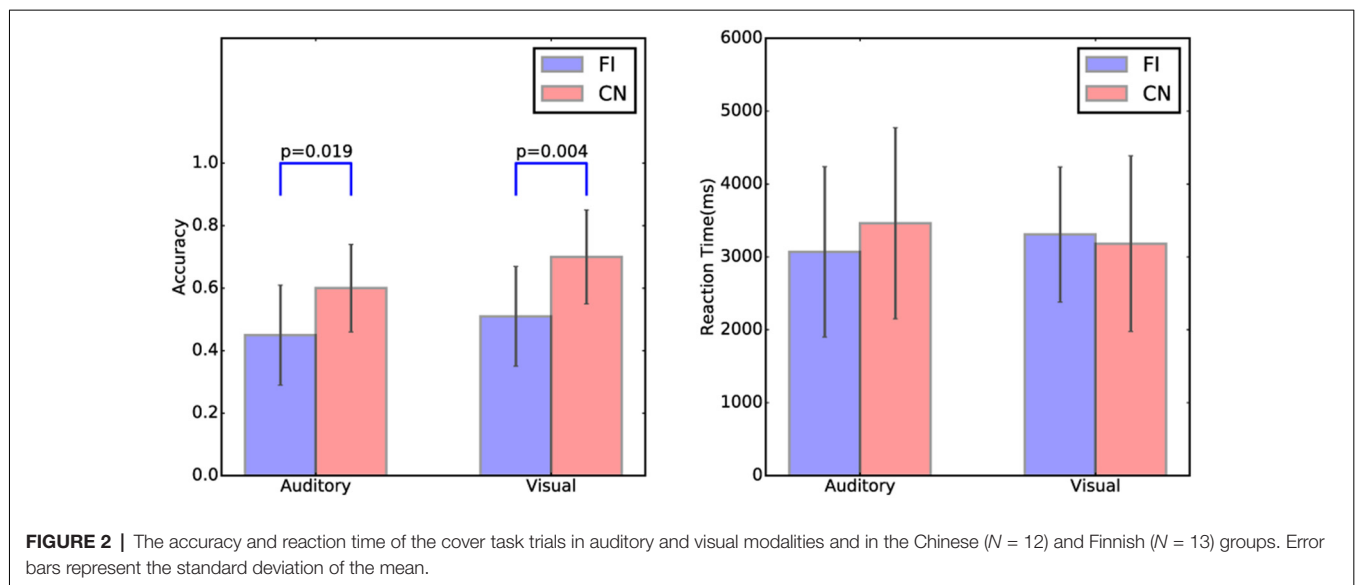
together (the numbers of trials in A and V conditions were equalized). To match the noise level of the A + V and AVC conditions and thus make these two conditions comparable, a subset of the AVC trials was created by randomly selecting about half the number of trials from the AVC condition as the number of trials in the A + V condition. For sensor level comparison, the gradiometer channel pairs in two orthogonal directions were combined using the vector sum method implemented in FieldTrip toolbox (Oostenveld et al., 2011). Gradiometers were chosen because they are less sensitive to noise sources originating far from the sensors than the magnetometers.

Individual magnetic resonance images (MRI) were not available from the participants and therefore the Freesurfer (RRID:SCR_001847) average brain (FSAverage) was used as a template for the source analysis. Three-parameter scaling was used to coregister FSAverage with the individual digitized head points. The average distance between the digitized head points and the scaled template scalp surface was 3.55 mm (0.25 mm SD).

Depth-weighted L2-minimum-norm estimate (wMNE; Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006) was calculated for 10,242 free orientation current dipoles distribute on the cortical surface in each hemisphere. The noise covariance matrix was estimated from the raw MEG data from the 200 ms pre-stimulus baseline. The inverse solution was noise-normalized using dynamical statistical parametric maps (dSPM; Dale et al., 2000) for further statistical analysis. Since the FSAverage brain template that was used for all participants was only scaled to the subject-specific head size, the estimated brain activity could be directly compared in the statistical analyses without morphing to a common brain.

Statistical Analysis

Sensor level statistical analyses on the combined gradiometers were conducted using the nonparametric permutation test (Maris and Oostenveld, 2007) with spatial-temporal clustering based on t-test statistic implemented in the FieldTrip toolbox.

FIGURE 2 | The accuracy and reaction time of the cover task trials in auditory and visual modalities and in the Chinese (N = 12) and Finnish (N = 13) groups. Error bars represent the standard deviation of the mean.

Source level analyses were similarly conducted using the nonparametric permutation t-test with spatio-temporal clustering in MNE Python. The time window was selected from 0 to 1,000 ms after the stimulus onset and the number of permutations was 1,024 in both sensor and source level statistical analyses. The cluster alpha was 0.05 for the incongruent–congruent comparison. For the AV − (A + V) comparison, a more conservative cluster alpha value of 0.005 was used due to the lower SNR since only half of the trials were averaged compared with the congruency contrast.

RESULTS

Behavioral Performance

The accuracy and reaction time of the cover task trials in both Chinese and Finnish groups are shown in **Figure 2**. There was a significant group difference in the accuracy for both auditory ($t_{(23)} = -2.52$, $p = 0.019$) and visual modality ($t_{(23)} = -3.20$, $p = 0.004$) between Chinese and Finnish groups, with the Chinese participants (Auditory: mean = 0.60, SD = 0.13 ; Visual: mean = 0.70, SD = 0.15) being more accurate than the

Finnish participants (Auditory: mean = 0.45, SD = 0.16; Visual: mean = 0.51, SD = 0.16) as expected. No significant differences in the reaction time were found between Chinese and Finnish groups.

Grand Average

Figure 3 gives an overview of the brain responses to the unimodal and bimodal audiovisual stimuli in sensor and source level in both Chinese and Finnish groups. **Figure 3A** shows the grand average evoked waveforms for the A, V, AVC and AVI conditions averaged over left and right temporal and occipital channels (vector sum of the paired orthogonal gradiometer channels) in the Chinese and Finnish groups. **Figure 3B** shows the magnetic field topography and dSPM source activations of the peak evoked responses at early (100–200 ms) and late (300–700 ms) time windows for each of the four conditions.

Suppression Effect [AVC − (A + V)]

Testing for suppression effects in the latency range from 0 to 1,000 ms post-stimulus in both groups separately at the sensor level, the cluster-based permutation test revealed significant
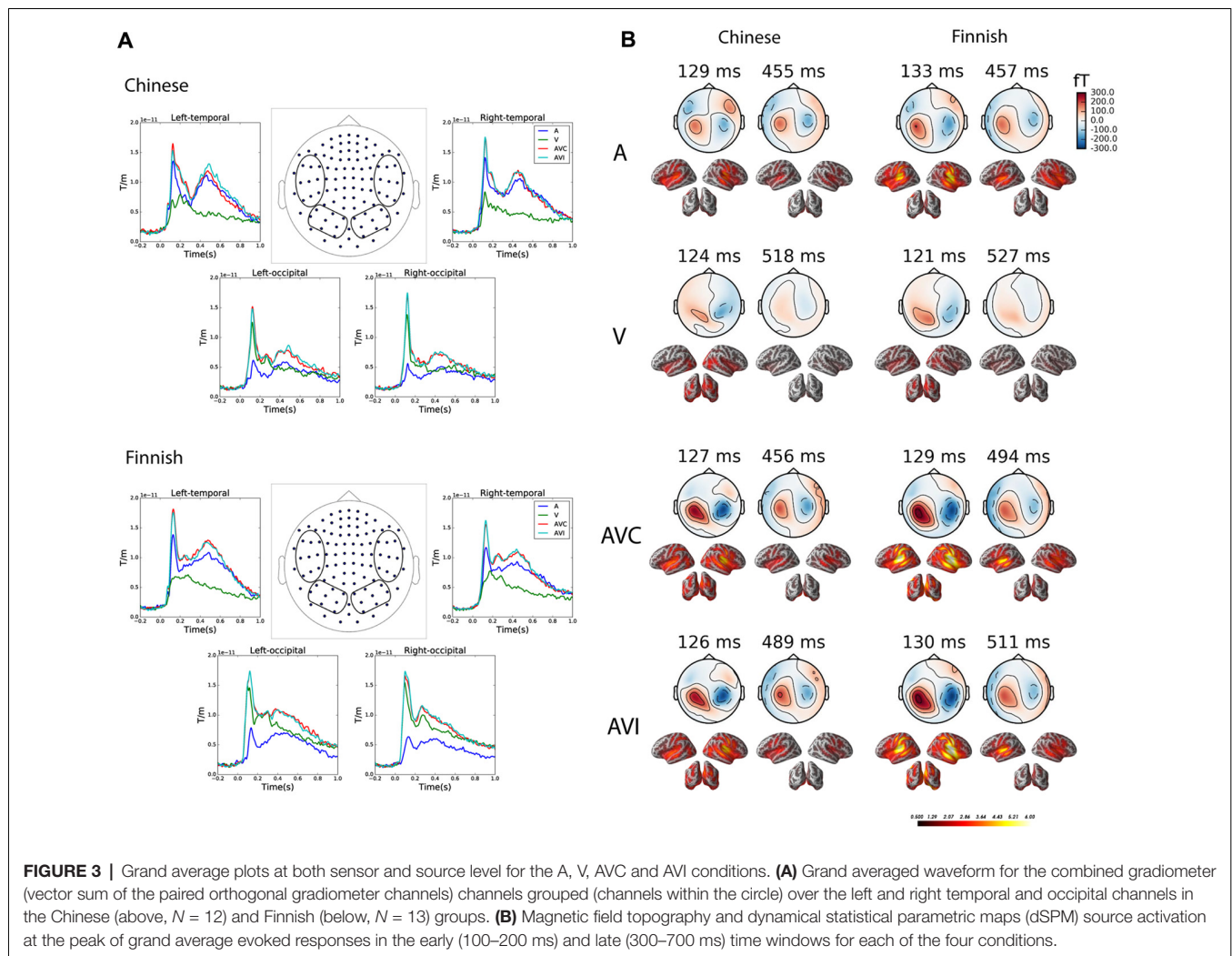

FIGURE 3 | Grand average plots at both sensor and source level for the A, V, AVC and AVI conditions. **(A)** Grand averaged waveform for the combined gradiometer (vector sum of the paired orthogonal gradiometer channels) channels grouped (channels within the circle) over the left and right temporal and occipital channels in the Chinese (above, $N = 12$) and Finnish (below, $N = 13$) groups. **(B)** Magnetic field topography and dynamical statistical parametric maps (dSPM) source activation at the peak of grand average evoked responses in the early (100–200 ms) and late (300–700 ms) time windows for each of the four conditions.

differences between the summed auditory and visual only conditions and the AVC conditions for both the Chinese group (p = 0.002) and the Finnish group (p = 0.006). In the Chinese group, the significant cluster was from 557 to 692 ms and mainly at the left temporal and frontal channels. In the Finnish group, the significant cluster was from 363 to 520 ms and mainly at the right parietal-occipital channels.

At the source level, the cluster-based permutation test also revealed significant differences between the summed auditory and visual only conditions and AVC conditions for both the Chinese group and the Finnish group. In the Chinese group, two significant clusters were found with different time windows and source locations. The first significant cluster was from 205 to 365 ms in the left angular and supramarginal gyri (p = 0.01). The second significant cluster was from 575 to 800 ms in the left temporal and frontal regions (p = 0.001). In the Finnish group, the significant cluster was from 285 to 460 ms and found in the right parietal-occipital regions (p = 0.003).

Results for both sensor and source level suppression integration comparisons and the spatiotemporal pattern of significant clusters are shown in **Figure 4**.

Congruency Effect (AVI − AVC)

At the sensor level, the congruency effect was tested in the latency range from 0 to 1,000 ms post-stimulus in both groups separately. The cluster-based permutation test revealed a significant difference between the congruent and incongruent conditions in the Chinese group only (p = 0.01; see **Figure 5**). The cluster occurred from 538 to 690 ms and the difference was most pronounced over the left frontal and temporal sensors in this latency range.

At the source level, the cluster-based permutation test also revealed a significant difference between the congruent and incongruent conditions in the Chinese group only (p = 0.008; see **Figure 5**). The cluster occurred from 490 to 890 ms and the difference was most pronounced over the left frontal and temporal regions.

DISCUSSION

In this study, we examined the effect of long-term exposure to Chinese characters on the audiovisual integration processes. We used a rigorous control of the participant's attention on the two modalities by introducing a dual-modality one-back task. We observed that exposure to one's native language and the long-term memory traces for spoken and written language indeed significantly changed the neural networks of audiovisual integration. This can be seen in the left-lateralized suppression effect, reflecting automatic audiovisual processes in the Chinese participants and a right-lateralized suppression effect in Finnish participants when processing novel Chinese audiovisual stimuli. Furthermore, long-term memory representations of Chinese characters are also manifested in how the brain reacts to the AVC and AVI stimuli in the left superior temporal and Broca's area in native Chinese speakers.
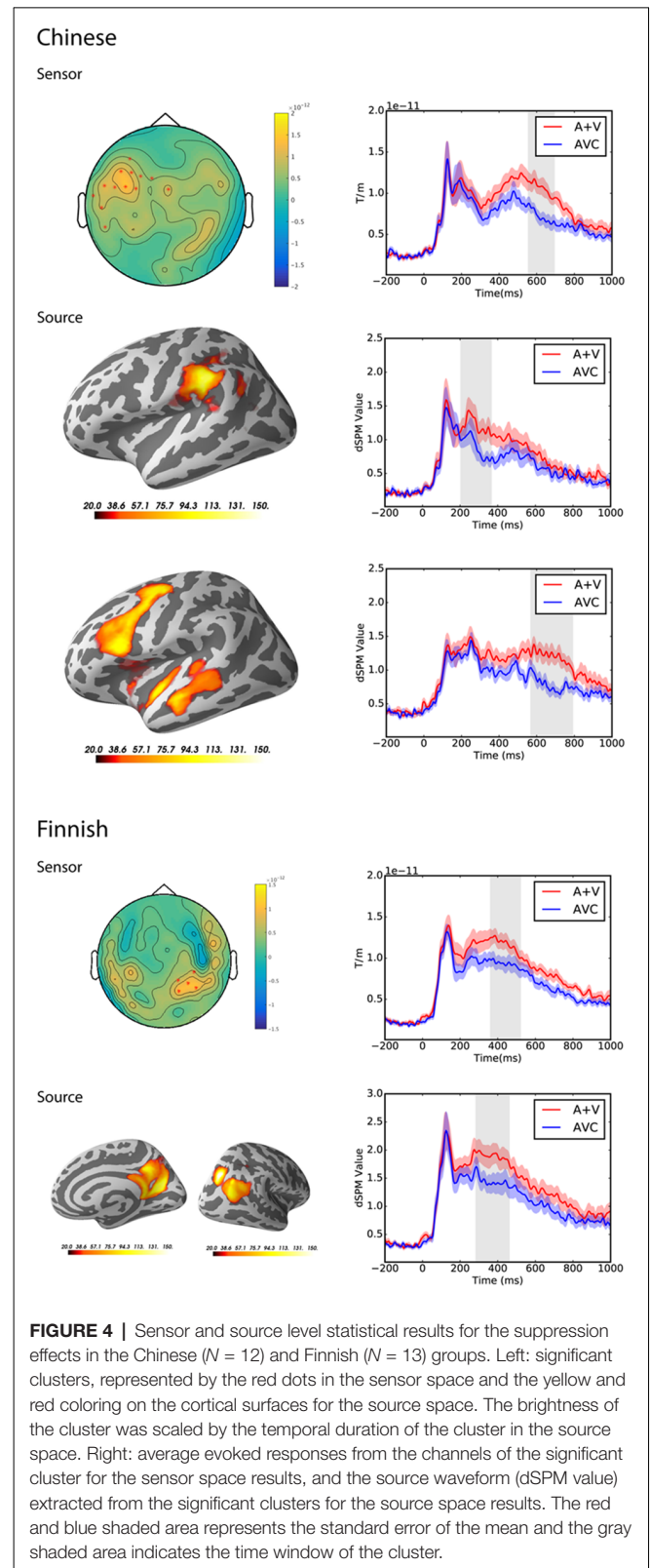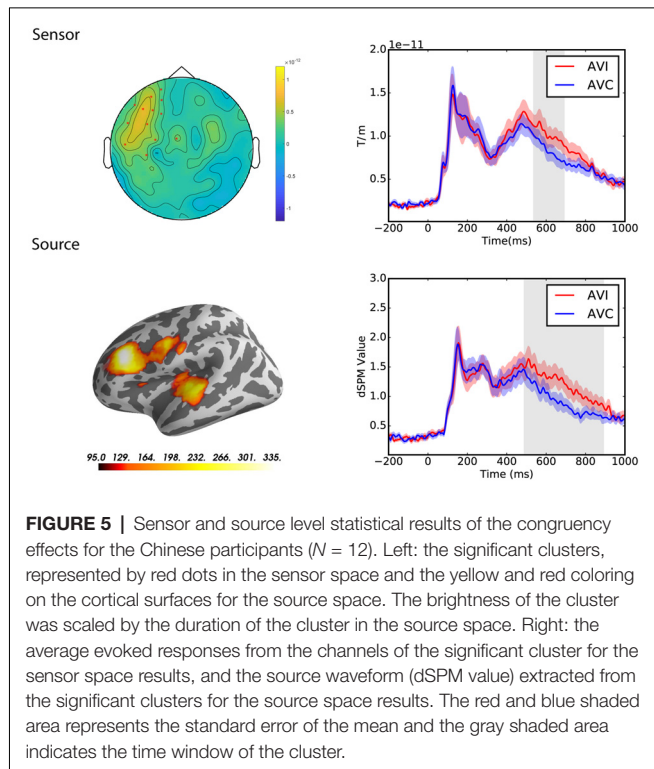

FIGURE 4 | Sensor and source level statistical results for the suppression effects in the Chinese (N = 12) and Finnish (N = 13) groups. Left: significant clusters, represented by the red dots in the sensor space and the yellow and red coloring on the cortical surfaces for the source space. The brightness of the cluster was scaled by the temporal duration of the cluster in the source space. Right: average evoked responses from the channels of the significant cluster for the sensor space results, and the source waveform (dSPM value) extracted from the significant clusters for the source space results. The red and blue shaded area represents the standard error of the mean and the gray shaded area indicates the time window of the cluster.

Suppression (AVC vs. A + V) effects were found in both the Chinese and Finnish groups but with different hemispheric lateralizations and at different time windows.

FIGURE 5 | Sensor and source level statistical results of the congruency effects for the Chinese participants (N = 12). Left: the significant clusters, represented by red dots in the sensor space and the yellow and red coloring on the cortical surfaces for the source space. The brightness of the cluster was scaled by the duration of the cluster in the source space. Right: the average evoked responses from the channels of the significant cluster for the sensor space results, and the source waveform (dSPM value) extracted from the significant clusters for the source space results. The red and blue shaded area represents the standard error of the mean and the gray shaded area indicates the time window of the cluster.

The suppression effect was most pronounced over the left angular/supramarginal, temporal and inferior frontal regions in the Chinese group, which suggested that the left hemisphere language network of native Chinese speakers was activated during processing of learned audiovisual association of Chinese characters and speech sounds. When children learn to read in an alphabetic language, visual letters and auditory phonemes are often presented simultaneously and neural pathways that map graphemes to phonemes are formed (Brem et al., 2010). Neural connectivity is strengthened through learning to read in school and later through reading practice, which would enable fluent readers to form long-term memory traces for written symbols and fast retrieval of audiovisual associations (Maurer et al., 2006). Consequently, the audiovisual suppression effect can be interpreted as the optimization of neural networks as a consequence of language learning in the Chinese participants (Raij et al., 2000). The early suppression effect was found in the left angular and supramarginal gyri about 200–350 ms after the stimulus onset. The left angular and supramarginal gyri have been considered as heteromodal areas related to linking orthographic representation from the occipital lobe to phonological coding represented in the STG (Price, 2000; Pugh et al., 2000; Schlaggar and McCandliss, 2007) and also possible feed forward connection to the inferior frontal gyrus (Simos et al., 2013). The left-lateralized suppression effect in the left superior and middle temporal areas in the later time window (about 550–800) matches those observed in earlier MEG, EEG and fMRI studies (Raij et al., 2000; Calvert et al., 2001; van Atteveldt et al., 2004) using alphabetic letters. The

effect in Broca's area is less often reported in audiovisual integration studies using alphabetic languages, possibly due to the fact that the Chinese language requires extra semantic processing in this area as shown by many other studies (Kuo et al., 2001; Tan et al., 2001; Wang et al., 2008). Taken together, it seems that the left superior temporal cortex is a common node in the neural network for audiovisual language processing for both alphabetic and logographic scripts. Importantly, the left inferior frontal cortex seems to be involved in the additional semantic-related audiovisual processing in logographic scripts (Wu et al., 2012; van Atteveldt and Ansari, 2014).

In the Finnish group, the suppression effect was most pronounced in the right inferior parietal and occipital area in a slightly earlier time window (285–460 ms), which partly matches with the findings by Raij et al. (2000) who also showed a predominant audiovisual interaction effect in the right temporo-occipito-parietal junction (280–345 ms). Clearly, for Finnish participants, Chinese characters did not have long-term memory representations and might be processed more like line drawings instead of meaningful characters. Parietal areas have been shown to be involved in multisensory processing in both human and monkey studies (Grunewald et al., 1999; Lewis et al., 2000; Bremmer et al., 2001). For example, lateral intraparietal area (LIP) which was originally considered as part of the unimodal visual cortex has been shown to respond to auditory stimuli in many studies (Bremmer et al., 2001; Ahveninen et al., 2006; Cohen, 2009). Other research has also indicated that the right parietal region plays a role in the perception of stimuli without any long-term representations arising simultaneously from the multiple sensory modalities (Kamke et al., 2012). The suppression effect in the right parietal and occipital regions at early time window indicates that the unfamiliar audiovisual information processing might rely more on the processing of the visual features (Calvert et al., 2001; Madec et al., 2016). It could be explained by the fact that Finnish participants who are naive to Chinese focused much more attention on the analysis of the spatial information of various strokes comprising the logographic character in order to be able to integrate the audiovisual stimuli. It should be noted that the suppression effect in the Finnish group was mostly related to this specific task and should not be interpreted as audiovisual integration related to logographic language. Therefore, the suppression effect found in the Finnish participants most likely represents a basic audiovisual processing (Calvert, 2001; Molholm et al., 2002; Cappe et al., 2010) of novel symbols in a working memory task.

Spatial-temporal clustering revealed an N400-like deflection elicited due to the congruency of the audiovisual stimuli in the Chinese, which was absent in the Finnish group. The congruency effect was mainly localized close to the left superior temporal cortex, Heschl's gyrus, inferior frontal cortex and also parts of the insula. The difference occurred in a late time window around 500–800 ms. This result is in line with previous studies using MEG and EEG (Raij et al., 2000; Jost et al., 2014). According to the functional neuroanatomical model for the integration of letters and

speech sounds proposed by van Atteveldt et al. (2009), the auditory and visual information is integrated in the heteromodal STS/STG and then feedback projected to the auditory cortex. In the current study, the congruency effect presented in the late time window therefore supports such feedback projection mechanism. Previous studies (Raij et al., 2000; van Atteveldt et al., 2004) using transparent alphabetic grapheme-phoneme associations found that the brain activations in the auditory association cortex were enhanced by congruent letters and suppressed by incongruent letters. This modulation is overruled during the explicit matching when both types of the audiovisual stimuli are equally relevant, independent of the congruency and temporal relation (van Atteveldt et al., 2007b). Whereas in opaque languages such as English, incongruent grapheme-phoneme associations have been shown to elicit a weaker and even reversed pattern of audiovisual integration to congruent pairs in the superior temporal areas (Holloway et al., 2015). Our results further indicated that audiovisual congruency is adaptive to different script types with a reversed direction in the superior temporal areas in logographic language compared to earlier reports on alphabetic language (Raij et al., 2000; van Atteveldt et al., 2004). Converging evidence from another audiovisual integration study (Jost et al., 2014) revealed a similar congruency effect for familiar German words, and was attributed to lexical-semantic processes involved in the processing of audiovisual words stimuli. The inferior frontal cortex has repeatedly been shown to be activated specifically by semantically incongruent audiovisual stimuli, which has been attributed to increased demands on the cognitive control involving semantic retrieval and working memory processes (Martin and Chao, 2001; Wagner et al., 2001; Doehrmann and Naumer, 2008).

The Chinese group showed a decrease of activation in the inferior frontal cortex when congruent audiovisual pairs were compared to the combined auditory and visual only stimulations. This indicates less neural resources for the audiovisual processing than the sum of resources it took to process the auditory and visual information independently. Furthermore, in the congruency comparison, the incongruent audiovisual pairs showed a higher activation than the congruent pairs again suggesting that it is less demanding to process the congruent audiovisual information due to complementary representations from both auditory and visual modalities. Since the Finnish participants had never learned the character-speech sound associations, they did not show any congruency effects. The left-lateralized suppression and congruency effects in the Chinese group suggest that the native Chinese could effectively utilize the audiovisual features of the learned language, whereas the right-lateralized suppression effect in the Finnish group may suggest that the Finnish participants rely more on the basic audiovisual processing mechanisms including substantial visual feature analysis. Therefore, congruency and suppression effects provide complementary information about audiovisual integration.

Both suppression and congruency effects in the Chinese group were mainly in the late time window about 500–800 ms after

the stimulus onset. Previous studies have also found similar late orthographic-phonological interactions (400–700 ms) in spoken language processing during metaphonological tasks (Pattamadilok et al., 2011; Lafontaine et al., 2012). Interestingly, a late negativity enhancement (about 650 ms) after auditory stimulus onset was reported in mismatch negativity (MMN) studies on audiovisual integration in both the beginner and advanced readers in children (Froyen et al., 2009; Zaric et al., 2014, 2015). It was interpreted as more elaborate associative processes that are activated for the integration of letter-speech pairs in children. In addition, the suppression and congruency effects were mainly localized in the left frontal and superior and middle temporal areas. These locations are consistent with previous fMRI studies using Chinese characters and have shown that character reading process is characterized by particularly strong left lateralization of the frontal (BAs 9 and 47) and temporal cortices (Tan et al., 2000, 2001). For example, the left middle frontal area (BA 9) was suggested as an important region for integrating the logographic visuospatial analysis and the semantic processing in Chinese (Tan et al., 2001). Given the fact that the characters/speech items all refer to meaningful words, both the suppression as well as the congruency effect in the late time window most likely reflect a mixture of modulatory feedback and lexico-semantic processing during audiovisual integration in the Chinese participants. However, as discussed earlier, such audiovisual modulatory feedback and lexico-semantic processing seem to be largely overlapping in time windows and brain regions, further studies are needed to dissociate the effects of semantic processing from the other audiovisual process.

There are some limitations in the present study. We used an active paradigm which corresponded to real life situations of audiovisual integration (e.g., learning to read) as an active action. However, it also brought some challenges to identify the underlying functional brain networks while the participants were performing a working memory task. The brain activity during an active task makes the interpretation of direct group comparisons problematic since the task difficulties and demands for the two groups were different. Therefore, the conclusion regarding the orthographic influences on multisensory integration was indirectly based on the comparison of our results in the Chinese group with previous letter-speech sound integration studies in alphabetic languages. Another limitation relates to the lack of structural magnetic resonance images in the present study, which could potentially lead to poorer source localization accuracy. However, the localization of the brain activity at the millimeter scale was not the goal of the current study, instead we were interested in the estimation of the larger cortical areas that contributed to the cognitive process.

CONCLUSION

Our findings demonstrated the effect of long-term exposure to logographic language on audiovisual integration processes

for written characters and speech sounds. Different suppression effects in the Chinese and Finnish groups indicated the adaptive nature of the brain networks for processing different types of audiovisual information. Importantly, in the Chinese group, the left superior temporal and inferior frontal cortices were actively involved in the processing of both audiovisual vs. unimodal and congruency information pinpointing the left superior temporal and frontal cortex as important hubs for audiovisual and semantic processing in Chinese. The findings are remarkably similar to those found for alphabetic languages in the left superior temporal cortex but with also unique aspects of the processing of the logographic stimuli in the inferior frontal cortex. The results thus indicate that there are universal audio-visual association mechanisms in language learning complemented by language-specific processes.

REFERENCES

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., et al. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proc. Natl. Acad. Sci. U S A* 103, 14608–14613. doi: 10.1073/pnas.0510480103

Andersen, T. S., Tiippana, K., and Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cogn. Brain Res.* 21, 301–308. doi: 10.1016/j.cogbrainres.2004.06.004

Bernstein, L. E., and Liebenthal, E. (2014). Neural pathways for visual speech perception. *Front. Neurosci.* 8:386. doi: 10.3389/fnins.2014.00386

Besle, J., Fort, A., and Giard, M.-H. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cogn. Process.* 5, 189–192. doi: 10.1007/s10339-004-0026-y

Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., and 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. doi: 10.1111/j.1460-9568.2008.06350.x

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., et al. (2010). Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U S A* 107, 7939–7944. doi: 10.1073/pnas.0904402107

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296. doi: 10.1016/s0896-6273(01)00198-2

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb. Cortex* 11, 1110–1123. doi: 10.1093/cercor/11.12.1110

Calvert, G. A., Hansen, P. C., Iversen, S. D., and Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage* 14, 427–438. doi: 10.1006/nimg.2001.0812

Calvert, G. A., and Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J. Physiol. Paris* 98, 191–205. doi: 10.1016/j.jphysparis.2004.03.018

Campanella, S., and Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn. Sci.* 11, 535–543. doi: 10.1016/j.tics.2007.10.001

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1001–1010. doi: 10.1098/rstb.2007.2155

Cappe, C., Thut, G., Romei, V., and Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580. doi: 10.1523/JNEUROSCI.1099-10.2010

Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hear. Res.* 258, 100–105. doi: 10.1016/j.heares.2009.01.011

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1

Doehrmann, O., and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res.* 1242, 136–150. doi: 10.1016/j.brainres.2008.03.071

Du, Y., Zhang, Q., and Zhang, J. X. (2014). Does N200 reflect semantic processing?—an ERP study on chinese visual word recognition. *PLoS One* 9:e90794. doi: 10.1371/journal.pone.0090794

Fort, A., Delpuech, C., Pernier, J., and Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb. Cortex* 12, 1031–1039. doi: 10.1093/cercor/12.10.1031

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn. Brain Res.* 10, 77–83. doi: 10.1016/s0926-6410(00)00024-0

Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., and Blomert, L. (2009). The long road to automation: neurocognitive development of letter-speech sound processing. *J. Cogn. Neurosci.* 21, 567–580. doi: 10.1162/jocn.2009.21061

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Grunewald, A., Linden, J. F., and Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I. effects of training. *J. Neurophysiol.* 82, 330–342. doi: 10.1152/jn.1999.82.1.330

Hämäläinen, M. S., and Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32, 35–42. doi: 10.1007/bf02512476

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J. Neurosci.* 27, 7881–7887. doi: 10.1523/JNEUROSCI.1740-07.2007

Herdman, A. T., Fujioka, T., Chau, W., Ross, B., Pantev, C., and Picton, T. W. (2006). Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neurosci. Lett.* 399, 61–66. doi: 10.1016/j.neulet.2006.01.069

Hocking, J., and Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain Lang.* 108, 89–96. doi: 10.1016/j.bandl.2008.10.005

Holloway, I. D., van Atteveldt, N., Blomert, L., and Ansari, D. (2015). Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. *Cereb. Cortex* 25, 1544–1553. doi: 10.1093/cercor/bht347

Jones, J. A., and Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14, 1129–1133. doi: 10.1097/00001756-200306110-00006

Jost, L. B., Eberhard-Moscicka, A. K., Frisch, C., Dellwo, V., and Maurer, U. (2014). Integration of spoken and written words in beginning readers: a topographic ERP study. *Brain Topogr.* 27, 786–800. doi: 10.1007/s10548-013-0336-4

Kamke, M. R., Vieth, H. E., Cottrell, D., and Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *Neuroimage* 62, 1334–1341. doi: 10.1016/j.neuroimage.2012.05.063

Kayser, C., Petkov, C. I., and Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cereb. Cortex* 18, 1560–1574. doi: 10.1093/cercor/bhm187

Kayser, S. J., Philiastides, M. G., and Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *Neuroimage* 148, 31–41. doi: 10.1016/j.neuroimage.2017.01.010

Kuo, W.-J., Yeh, T.-C., Duann, J.-R., Wu, Y.-T., Ho, L.-T., Hung, D., et al. (2001). A left-lateralized network for reading Chinese words: a 3 T fMRI study. *Neuroreport* 12, 3997–4001. doi: 10.1097/00001756-200112210-00029

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Lafontaine, H., Chetail, F., Colin, C., Kolinsky, R., and Pattamadilok, C. (2012). Role and activation time course of phonological and orthographic information during phoneme judgments. *Neuropsychologia* 50, 2897–2906. doi: 10.1016/j.neuropsychologia.2012.08.020

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., and Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp. Brain Res.* 166, 289–297. doi: 10.1007/s00221-005-2370-2

Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888. doi: 10.1093/cercor/10.9.873

Liberman, A. M. (1992). Chapter 9 the relation of speech to reading and writing. *Adv. Psychol.* 94, 167–178. doi: 10.1016/s0166-4115(08)62794-6

Lin, F. H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31, 160–171. doi: 10.1016/j.neuroimage.2005.11.054

Lütkenhöner, B., Lammertmann, C., Simões, C., and Hari, R. (2002). Magnetoencephalographic correlates of audiotactile interaction. *Neuroimage* 15, 509–522. doi: 10.1006/nimg.2001.0991

Madec, S., Le Goff, K., Anton, J.-L., Longcamp, M., Velay, J.-L., Nazarian, B., et al. (2016). Brain correlates of phonological recoding of visual symbols. *Neuroimage* 132, 359–372. doi: 10.1016/j.neuroimage.2016.02.010

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Martin, A., and Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Curr. Opin. Neurobiol.* 11, 194–201. doi: 10.1016/s0959-4388(00)00196-3

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., et al. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage* 33, 749–758. doi: 10.1016/j.neuroimage.2006.06.025

McBride, C. A. (2015). Is chinese special? Four aspects of chinese literacy acquisition that might distinguish learning chinese from learning alphabetic orthographies. *Educ. Psychol. Rev.* 28, 523–549. doi: 10.1007/s10648-015-9318-2

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cogn. Brain Res.* 14, 31–40. doi: 10.1016/s0926-6410(02)00059-9

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cogn. Brain Res.* 14, 115–128. doi: 10.1016/s0926-6410(02)00066-6

Murray, M. M., Lewkowicz, D. J., Amedi, A., and Wallace, M. T. (2016a). Multisensory processes: a balancing act across the lifespan. *Trends Neurosci.* 39, 567–579. doi: 10.1016/j.tins.2016.05.003

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., and Matusz, P. J. (2016b). The multisensory function of the human primary visual cortex. *Neuropsychologia* 83, 161–169. doi: 10.1016/j.neuropsychologia.2015.08.011

Murray, M. M., and Spierer, L. (2009). Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hear. Res.* 258, 121–133. doi: 10.1016/j.heares.2009.04.022

Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage* 25, 333–338. doi: 10.1016/j.neuroimage.2004.12.001

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Pattamadilok, C., Perre, L., and Ziegler, J. C. (2011). Beyond rhyme or reason: ERPs reveal task-specific activation of orthography on spoken language. *Brain Lang.* 116, 116–124. doi: 10.1016/j.bandl.2010.12.002

Perfetti, C. A., and Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 101–118. doi: 10.1037/0278-7393.24.1.101

Perrault, T. J. Jr., Vaughan, J. W., Stein, B. E., and Wallace, M. T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* 93, 2575–2586. doi: 10.1152/jn.00926.2004

Price, C. J. (2000). The anatomy of language: contributions from functional neuroimaging. *J. Anat.* 197, 335–359. doi: 10.1046/j.1469-7580.2000.19730335.x

Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., et al. (2000). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol. Sci.* 11, 51–56. doi: 10.1111/1467-9280.00214

Raij, T., Uutela, K., and Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron* 28, 617–625. doi: 10.1016/s0896-6273(00)00138-0

Rüsseler, J., Ye, Z., Gerth, I., Szycik, G. R., and Münte, T. F. (2018). Audio-visual speech perception in adult readers with dyslexia: an fMRI study. *Brain Imaging Behav.* 12, 357–368. doi: 10.1007/s11682-017-9694-y

Schlaggar, B. L., and McCandliss, B. D. (2007). Development of neural systems for reading. *Annu. Rev. Neurosci.* 30, 475–503. doi: 10.1146/annurev.neuro.28.061604.135645

Schroeder, C. E., and Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn. Brain Res.* 14, 187–198. doi: 10.1016/s0926-6410(02)00073-3

Schröger, E., and Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* 35, 755–759. doi: 10.1111/1469-8986.3560755

Service, E., Helenius, P., Maury, S., and Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *J. Cogn. Neurosci.* 19, 1193–1205. doi: 10.1162/jocn.2007.19.7.1193

Shu, H. (2003). Chinese writing system and learning to read. *Int. J. Psychol.* 38, 274–285. doi: 10.1080/00207590344000060

Simos, P. G., Rezaie, R., Fletcher, J. M., and Papanicolaou, A. C. (2013). Time-constrained functional connectivity analysis of cortical networks underlying phonological decoding in typically developing school-aged children: a magnetoencephalography study. *Brain Lang.* 125, 156–164. doi: 10.1016/j.bandl.2012.07.003

Sperdin, H. F., Cappe, C., Foxe, J. J., and Murray, M. M. (2009). Early, low-level auditory-somatosensory multisensory interactions impact reaction time speed. *Front. Integr. Neurosci.* 3:2. doi: 10.3389/neuro.07.002.2009

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331

Tan, L. H., Liu, H. L., Perfetti, C. A., Spinks, J. A., Fox, P. T., and Gao, J. H. (2001). The neural system underlying Chinese logograph reading. *Neuroimage* 13, 836–846. doi: 10.1006/nimg.2001.0749

Tan, L. H., Spinks, J. A., Eden, G. F., Perfetti, C. A., and Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proc. Natl. Acad. Sci. U S A* 102, 8781–8785. doi: 10.1073/pnas.0503523102

Tan, L. H., Spinks, J. A., Gao, J.-H., Liu, H.-L., Perfetti, C. A., Xiong, J., et al. (2000). Brain activation in the processing of Chinese characters and words: a functional MRI study. *Hum. Brain Mapp.* 10, 16–27. doi: 10.1002/(sici)1097-0193(200005)10:1<16::aid-hbm30>3.0.co;2-m

Taulu, S., and Kajola, M. (2005). Presentation of electromagnetic multichannel data: the signal space separation method. *J. Appl. Phys.* 97:124905. doi: 10.1063/1.1935742

Taulu, S., Kajola, M., and Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topogr.* 16, 269–275. doi: 10.1023/b:brat.0000032864.93890.f9

Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51, 1759–1768. doi: 10.1088/0031-9155/51/7/008

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., and Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn. Brain Res.* 14, 106–114. doi: 10.1016/s0926-6410(02)00065-4

van Atteveldt, N., and Ansari, D. (2014). How symbols transform brain function: a review in memory of Leo Blomert. *Trends Neurosci. Edu.* 3, 44–49. doi: 10.1016/j.tine.2014.04.001

van Atteveldt, N. M., Formisano, E., Blomert, L., and Goebel, R. (2007a). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb. Cortex* 17, 962–974. doi: 10.1093/cercor/bhl007

van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007b). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065

van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025

van Atteveldt, N., Murray, M. M., Thut, G., and Schroeder, C. E. (2014). Multisensory integration: flexible use of general operations. *Neuron* 81, 1240–1253. doi: 10.1016/j.neuron.2014.02.044

van Atteveldt, N., Roebroeck, A., and Goebel, R. (2009). Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear. Res.* 258, 152–164. doi: 10.1016/j.heares.2009.05.007

Vartiainen, J., Parviainen, T., and Salmelin, R. (2009). Spatiotemporal convergence of semantic processing in reading and speech perception. *J. Neurosci.* 29, 9271–9280. doi: 10.1523/JNEUROSCI.5860-08.2009

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338. doi: 10.1016/S0896-6273(01)00359-2

Wang, S., Zhu, Z., Zhang, J. X., Wang, Z., Xiao, Z., Xiang, H., et al. (2008). Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia* 46, 1371–1378. doi: 10.1016/j.neuropsychologia.2007.12.020

Wu, C.-Y., Ho, M.-H. R., and Chen, S.-H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage* 63, 381–391. doi: 10.1016/j.neuroimage.2012.06.047

Xu, W., Kolozsvari, O. B., Monto, S. P., and Hämäläinen, J. A. (2018). Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Front. Hum. Neurosci.* 12:304. doi: 10.3389/fnhum.2018.00304

Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., and Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One* 9:e110337. doi: 10.1371/journal.pone.0110337

Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., and Bonte, M. (2015). Crossmodal deficit in dyslexic children: practice affects the neural timing of letter-speech sound integration. *Front. Hum. Neurosci.* 9:369. doi: 10.3389/fnhum.2015.00369

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3