SCISPACE
formerly Typeset

Open access • Posted Content • DOI:10.1101/2020.05.31.126128

Neurofunctional mechanisms underlying audiovisual integration of characters and pinyin in Chinese children — Source link

Zhichao Xia, Zhichao Xia, Ting Yang, Cui X ...+5 more authors

Institutions: Beijing Normal University, McGovern Institute for Brain Research, University of Connecticut

Published on: 01 Jun 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Orthographic depth, Pinyin, Temporal cortex and Orthography

Related papers:

- Neurofunctional basis underlying audiovisual integration of print and speech sound in Chinese children
- Orthographic dependency in the neural correlates of reading: Evidence from audiovisual integration in English readers
- Neural Correlates of Morphological Processing: Evidence from Chinese.
- Hemispheric processing of lexical information in Chinese character recognition and its relationship to reading performance.
- Functional and anatomical dissociation between the orthographic lexicon and the orthographic buffer revealed in reading and writing Chinese characters by fMRI.

Share this paper:

View more about this paper here: https://typeset.io/papers/neurofunctional-mechanisms-underlying-audiovisual-1cfcga2qk6

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Neurofunctional mechanisms underlying audiovisual integration of characters and *pinyin* in Chinese children

Zhichao Xia [a, b *], Ting Yang [c, d *], Xin Cui [a], Fumiko Hoeft [e, f, g, h], Hong Liu [c, d, e], Hua Shu [a †], Xiangping Liu [c, d †]

[a] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, China

[b] School of Systems Science, Beijing Normal University, China

[c] Faculty of Psychology, Beijing Normal University, China

[d] Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, China

[e] Department of Psychological Sciences and Brain Imaging Research Center, University of Connecticut, USA

[f] Department of Psychiatry and Weill Institute for Neurosciences and Dyslexia Center, University of California, San Francisco, USA

[g] Haskins Laboratories, USA

[h] Department of Neuropsychiatry, Keio University School of Medicine, Japan

[*] These authors contributed equally to this work

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

† Corresponding authors: Dr. Hua Shu, State Key Lab of Cognitive Neuroscience and Learning, Beijing Normal University, China. Email: shuhua@bnu.edu.cn; Dr. Xiangping Liu, Faculty of Psychology, Beijing Normal University, China. Email: lxp599@163.com

Running Title: Brain basis of audiovisual integration

Number of pages: 46

Number of words in Abstract: 250, Introduction: 1053, Discussion: 1974.

Number of Tables: 2, Figures: 4

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Efficient integration of grapheme and phoneme is crucial for reading development in alphabetic languages, and superior temporal cortex (STC) has been demonstrated to be the most critical region for this process. To determine whether a similar neural mechanism underlies such processing in non-alphabetic languages, and to explore its relationship with reading, we conducted a functional magnetic resonance imaging study in typically developing Chinese children. Highly frequent pictographic characters and *pinyin*, a transparent alphabetic coding system that assists individuals to learn the pronunciation of new characters were investigated. In support of the orthographic depth hypothesis developed in alphabetic languages, reverse congruency effect (i.e., higher activation in incongruent condition compared to congruent condition) was identified in the left inferior frontal gyrus (IFG) and bilateral STC for character (deep orthography) conditions. Furthermore, correlation analysis revealed that the congruency contrast in the left IFG was associated with proficiency in reading comprehension and morphological awareness, suggesting reading-related semantic access during the implicit integration of multisensory information regarding characters. For *pinyin* (shallow orthography) conditions, while no regions showed a significant congruency effect at the group level in either direction, the congruency contrast in the left superior temporal gyrus was positively correlated with oral reading performance. This observation is consistent with findings in transparent scripts on reading disorders. Taken together, these results support and generalize the orthographic depth hypothesis to a logographic

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

language, highlight the role of the left IFG in character-syllable integration, and underscore the role *pinyin* could play in developing fluent reading in Chinese children.

Keywords

Audiovisual integration, character, fMRI, *pinyin*, Chinese reading development

Highlights

- The left IFG and bilateral STC are involved in audiovisual integration of characters.
- The congruency contrast in the left IFG in response to characters corresponds to semantic access.
- The congruency contrast in the left STG in response to *pinyin* is correlated with print-to-sound mapping.
- These findings support the orthographic depth hypothesis in a nonalphabetic language.
- The results imply a close link between *pinyin* and Chinese reading acquisition.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

Audiovisual integration of print and speech sound is a prerequisite for successful reading acquisition (Blomert & Froyen, 2010; Richlan, 2019). Behavioral and neurofunctional evidence has revealed that the lack of automatic print-to-sound integration uniquely accounts for the reading failure in shallow orthographies beyond commonly reported reading-related cognitive-linguistic skills, such as phonological awareness (PA) (Bakos, Landerl, Bartling, Schulte-Körne, & Moll, 2017; Blau et al., 2010; Leo Blomert & Willems, 2010).

Better knowledge of the neural basis underlying audiovisual integration can further our understanding of how this process contributes to (a)typical reading development and provide objective training targets and evaluation indices for children with reading disorders (RD). To achieve this goal, neuroimaging techniques, especially functional magnetic resonance imaging (fMRI), have been widely applied in the past two decades. Across these studies, distinguishable brain responses to congruent versus incongruent audiovisual stimulus pairs have been demonstrated to be the most reliable measure of multisensory integration and have been named the 'congruency effect' (Blau et al., 2010; Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009; Holloway, van Atteveldt, Blomert, & Ansari, 2015; van Atteveldt, Formisano, Goebel, & Blomert, 2004; van Atteveldt, Formisano, Goebel, & Blomert, 2007). For example, in a pioneering study by van Atteveldt et al. (2004), superior temporal cortex (STC) showed a stronger activation in congruent than in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incongruent conditions. Furthermore, while the involvement of STC in the audiovisual integration of print and speech sound has been replicated regardless of the level of processing (e.g., phoneme, syllable, or whole word) (Kast, Bezzola, Jancke, & Meyer, 2011; Mcnorgan, Randazzo-Wagner, & Booth, 2013), researchers also found that the direction of the effect and the area displaying such effect are influenced by multiple factors, where orthographic depth was an important one (Holloway et al., 2015; Xu, Kolozsvari, Monto, & Hämäläinen, 2018).

Orthographic depth refers to the consistency of mapping between visual and phonological forms in a writing system (e.g., grapheme-phoneme correspondence in alphabetic languages, GPC hereafter). The orthographic depth hypothesis would suggest that the congruency effect during audiovisual integration is in a direction that is associated with the consistency of GPC. In support of this, researchers have found that native speakers of English (deep orthography) showed a reverse congruency effect (aka incongruency effect, higher activation for incongruent condition), whereas there is a congruency effect (higher activation for congruent condition) observed in Dutch (shallow orthography) (Holloway et al., 2015). More importantly, when the same English-speaking participants processed digits, where there is one-to-one mapping, the effect became positive. Although the conclusion is clear, these findings are restricted to alphabetic languages. In order to broaden the applicability of this hypothesis, research in non-alphabetic languages is necessary.

Chinese is particularly suited to for answer this question, given its logographic nature and extremely deep orthography. In general, character is the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

basic graphic unit. While different characters may share the same pronunciation (i.e., homophones), a single character can also have multiple sounds (i.e., polyphone). Therefore, the link between print (character) and speech sound (syllable) is quite opaque. According to the orthographic depth hypothesis, a reverse congruency effect should be observed. Moreover, since grapho-semantic correspondence is more systematic in Chinese and plays a role equal to or even more important than phonological processing in Chinese reading acquisition (Liu et al., 2017; Ruan, Georgiou, Song, Li, & Shu, 2018; Yang, Shu, McCandliss, & Zevin, 2012), additional regions such as the inferior and middle frontal cortex known to be involved in semantic processing may be recruited (Wu, Ho, & Chen, 2012; Zhao et al., 2014). Recently, Xu et al., (2019) conducted a magnetoencephalography (MEG) study with an one-back task in Chinese adults and identified a reverse congruency effect across the left frontal cortex and STC. However, some limitations need to be considered, including small sample size (n = 12), using flat tone speech sounds that do not exactly correspond to any character, and source reconstruction accuracy of MEG. Moreover, in experienced readers, audiovisual integration of characters and corresponding sound is fully automated. Thus, in addition to address these issues, an important follow-up question should be neurofunctional features of this process in developing populations. Against this background, the first aim of the current study is to determine whether there is a cross-modal process supported by STC as well as other brain areas associated with integrating characters and speech sounds in typically developing Chinese children.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Of note, learning to read in Chinese is a longer process than learning most alphabetic languages (Shu, Chen, Anderson, Wu, & Xuan, 2003). At the very beginning stage, a small set of pictographic characters that are used as radicals in composite characters are taught first, since more than 80% of Chinese characters are phonograms that comprise semantic and phonetic radicals (Chan & Siegel, 2001). Given its important role, in this study, we specifically focused on these early-acquired single-element characters. Importantly, because pictographic characters do not contain radicals as phonological/semantic cues, the mapping between visual and sound/meaning is direct and completely arbitrary. Moreover, Chinese offers a valuable opportunity to examine the orthographic depth hypothesis using two different scripts in the same brain. *Pinyin* is a Roman alphabetic system used to represent the sound of characters in Mainland China and Singapore. A child will learn *pinyin* first and then use it to assist in conquering the pronunciation of new characters (Guan, Liu, Chan, Ye, & Perfetti, 2011; Yan, Miller, Li, & Shu, 2008). In other words, it is more like a scaffold during reading acquisition. Since *pinyin* borrows letters from English and has one-to-one mappings with speech sounds (phonemes), it is reasonable to predict a similar activation pattern in response to audiovisual stimulus pairs in *pinyin* and in transparent orthographies. Therefore, we included *pinyin* in this study additionally to further examine the orthographic depth hypothesis.

Finally, in addition to examining the congruency effect at the group level, understanding multisensory integration in relation to individual heterogeneity in

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reading is important. However, until now, the majority of research has been done by comparing children with RD with their typical peers (e.g., Blau et al., 2010; Blau et al., 2009; Plewko et al., 2018), leaving the brain-reading relationship in typically developing children underexplored. Therefore, we additionally adopted an individual differences approach to evaluate reading abilities and cognitive-linguistic skills in an effect to shed light on whether and how neural circuitries link multisensory integration of character/*pinyin* with typical reading acquisition.

2. Materials and Methods

2.1. Participants

Thirty children (20 girls; age 111 ~ 140 months, $M = 127$) were included in this study. They were selected from the initial sample of 55 typical children according to the following inclusion criteria: (a) native speaker of Chinese, (b) right-handed (Oldfield, 1971), (c) normal audition and normal or corrected-to-normal vision, (d) full-scale intelligence quotient (IQ) ≥ 85, (e) no history of neurological or psychiatric disorders, (f) typical reading (no less than 1 *SD* below the norm average in the reading screening task *Character Recognition*) (Xue, Shu, Li, Li, & Tian, 2013), (g) completed all four task scans, (h) accurate response rate (ACC) for each run > 75%, (i) no more than 17 (25%) 'bad' volumes in each run (see fMRI preprocessing). These children learned *pinyin* formally in the first half of the semester in the 1st grade, before learning characters. All the children and their parents/guardians were informed about the aim and content of the study and signed written consent before

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the experiment was conducted. All the behavioral and neuroimaging data were collected in 2018. This study received ethical approval from the Institutional Review Board of State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

2.2. Behavior measures

Each child received testing individually in a silent booth. IQ was measured by the abbreviated version of the *Chinese Wechsler Intelligence Scale for Children* (*WISC-CR*) (Wechsler, 1974): *Information*, *Similarities*, and *Digit Span* subtests were used to estimate verbal IQ and *Picture Completion*, *Block Design*, and *Coding* subtests were used to evaluate performance IQ.

A set of neuropsychological tasks was used to measure children's reading and reading-related cognitive-linguistic skills. In brief, as *Character Recognition* has been used in previous studies for identifying Chinese dyslexic children (Cui, Xia, Su, Shu, & Gong, 2016; Xia, Hoeft, Zhang, & Shu, 2016), we used it here for the screening aim. *Word List Reading* and *Silent Reading Comprehension* were used to measure oral reading fluency and reading comprehension proficiency, respectively. The former task relies more on grapho-to-phonological access, while the latter relies more on grapho-to-semantic processing (Xia et al., 2018). Lastly, tasks *Phoneme Deletion*, *Digit RAN*, and *Morphological Production* tasks were used to measure PA, RAN, and morphological awareness (MA), which are the most important cognitive-linguistic skills for reading development in Chinese (Lei et al., 2011).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Character Recognition is a standardized test for estimating the number of characters that the child has learned. It consists of 150 characters selected from textbooks for grades 1-6 (Xue et al., 2013). The characters are arranged in the order according to the grade on 10 sheets of A4 paper. During the task, the participant is asked to name the characters in sequence until they fail in all 15 items on one page. Each correct answer is worth 1 point, and the full mark is 150. No time limit was applied.

Word List Reading consists of 180 high-frequency two-character words (Zhang et al., 2012). These words are arranged in a 9-column by 20-row matrix in a sheet of A4 paper. The participant is required to read words aloud in a left-to-right, up-to-down order as accurately and quickly as possible. The time used to complete the task is recorded, and an index referring to the number of words the participant read correctly per minute is calculated as the final score.

Silent Reading Comprehension includes 100 sentences/paragraphs (Lei et al., 2011). The items are arranged in order based on length. The participant is requested to read each sentence/paragraph silently and decide the correctness of the meaning with mark of ✓ or ✗ as fast as possible within 3 minutes. The sum of the characters in sentences with correct responses is calculated. The final score indicates the number of characters the participant read in one minute.

Phoneme Deletion is used to assess PA (Li, Shu, McBride-Chang, Liu, & Peng, 2012). This task includes three sections, and the participant is asked to delete

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the initial, middle, or final sound from orally presented syllables. Each section consists of 3 practice trials and 8-10 testing items. In particular, the participant is asked to pronounce what is left when a phoneme (consonant/vowel) is deleted from a given syllable. One correct response is worth 1 point, and the total possible score is 28.

Digit RAN requires the participant to name 50 one-digit numbers on a card as accurately and rapidly as possible (Liu et al., 2017). Numbers 1, 2, 3, 5, and 8 repeat five times and arrange randomly in a 5-column by 10-row matrix. The task is administrated twice, and the average time the participant completes the task is used as the final score.

Morphological Production is used to measure MA at the character level (Shu, McBride-Chang, Wu, & Liu, 2006). In each trial, a character is orally presented in a word to the participant. The participant has to respond with two new words. In one word the given character retains the same meaning as that in the original case, while in the other word the meaning of the character should be different. This task contains 15 characters, with a full mark of 30.

2.3. Stimuli and experimental design

In this study, we investigated neurofunctional correlates underlying audiovisual integration in both characters and *pinyin*. In the character conditions, the stimuli were visually presented characters and their speech sounds (**Figure 1A**). Fifty-six highly frequent pictographic characters were selected (Chinese Single-Character

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Word Database; https://blclab.org/pyscholinguistic-norms-database/). All the characters are simple and taught early. In the experiment, visual stimuli were white characters displayed in 'KaiTi' font and 96 pt at the center of a black background. The speech sounds were recorded from a native Chinese male with a sampling rate of 44.1 kHz and 16-bit quantization. The audio files were normalized to 85 dB and then edited with a bandpass (100-4000 Hz) filter with Audacity (https://www.audacityteam.org/). The average duration of all sounds is 476.3 (±87.5) ms, without pre- or post-silence. All the speech sounds can be correctly recognized easily. For the *pinyin* conditions, the visual stimuli were the *pinyin* spelling of the characters (in white, 'Century Schoolbook' font, and 90 pt), while the auditory stimuli were the same as character conditions (**Figure 1B**).

The fMRI procedure was adapted from Blau et al. (2010) (**Figure 1C**). A block design was used with four conditions in each type of material: auditory-only (Aud), visual-only (Vis), audiovisual congruent (avC) and audiovisual incongruent (avI). The visual and auditory stimuli were presented simultaneously in the bimodal conditions. The entire experiment contained 4 functional runs, while the character and *pinyin* stimuli were presented in separate ones. To prevent character stimuli driving lexical access in *pinyin* processing, *pinyin* experiment was always conducted first. Each run included 8 experimental blocks (duration 20.8 s) with each condition replicated twice and interleaved with 9 rest blocks (duration 20.8 s). In each experimental block, there were 4 miniblocks. Each miniblock consisted of a 1.5-s brain volume acquisition and a 3.7-s silence (i.e., delay, see image acquisition

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

part) during which 4 trials of stimuli were presented. The order of the stimuli and conditions were pseudorandomized. Considering the age of participants, an irrelevant perceptual cover task was used to have children maintaining their attention on stimuli during scanning. Compared to the active matching task, a simple task unrelated to the process of interest will not change the automatic response pattern during the audiovisual integration (Blau et al., 2010; van Atteveldt et al., 2007). The participant was asked to press the button with the index finger of right hand accurately and quickly whenever the auditory (440 Hz pure tone) and/or visual (an unpronounceable symbol) targets appeared (**Figure 1D**). Only the one with an accurate rate equal to or above 75% in each run was included in the present study. To ensure that the child understood the task correctly, two practice sessions were administered, one outside the scanner and the other just before the first fMRI session inside the scanner.

2.4. Image acquisition

A 3-Tesla Siemens MAGNETOM Trio Tim scanner in Imaging Center for Brain Research at Beijing Normal University was used to collect all the images with a 12-channel head coil. In each child, 4 $T2^*$ sensitive gradient echo planar imaging sequences were collected with the following parameters: repetition time (TR) = 5200 ms, echo time = 32 ms, delay in TR = 3700 ms, flip angle = 90 degrees, slice thickness = 4.5 mm, voxel size = 3.0 × 3.0 × 4.5 mm^3, interscan gap = 0.675 mm, number of slices = 24, number of volumes = 68, and time of acquisition = 5 minutes and 54 s. Of note, this sparse sampling method provides a 3.7-s period of silence for

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

stimuli presentation. Such a design is recommended for fMRI experiments on the auditory cortex (Shah et al., 2000). In addition, a high-resolution whole-brain T1-weighted structural image (magnetization-prepared rapid acquisition with gradient echo, TR = 2530 ms, echo time = 3.39 ms, inversion time = 1100 ms, flip angle = 7 degrees, slice thickness = 1.33 mm, voxel size = 1.33 × 1.0 × 1.33 mm^3, number of axial slices = 144, and time of acquisition = 8 minutes and 7 s) was collected. Prior to the formal scan, children were informed about the experimental procedure, familiarized with the noise of MRI sequences, and trained to hold still. Foam pads were used to help reduce head motion during scanning. A radiologist blind to the details of the study reviewed all the images to ascertain any pathological deviations.

2.5. fMRI preprocessing

Preprocessing of images was conducted with SPM12 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl), including: (a) discarding the first 4 volumes to avoid T1 equilibration effect, (b) head motion correction (realign), (c) slice-time correction, (d) ART-based outlier detection (a 'bad' brain volume was defined as intensity > 9 SD or frame-to-frame head motion > 2 mm; https://www.nitrc.org/projects/artifact_detect), (e) T1 image segmentation, (f) normalization to the standard template in MNI (Montreal Neurological Institute) space, and (g) smoothing with an 8-mm full-width at half-maximum Gaussian kernel. In the subject level, each condition was modeled with a generalized linear model. Parameters referring to 'bad' volumes and head motion were included in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

model to remove the effect of nuisance covariates. Then, we calculated brain map of the congruency contrast (avC-avI), which was used in the subsequent analyses.

2.6. Statistical analyses

Descriptive statistics of demographic and behavioral measures were conducted first, followed by correlations between reading and cognitive-linguistic skills. Then, we evaluated in-scanner task performance. In this study, we defined an accurate response as a button press to the target(s) with a reaction time (RT) between 200 and 2000 ms. The average reaction time was calculated for correct responses. ACC and RT were compared across runs.

In the brain analysis, to identify regions showing congruency effect in character processing, one sample *t*-tests were conducted with individual contrast maps of congruency (i.e., congruent minus incongruent conditions). To avoid results produced by differences in de-activation, the neurobiological basis of which remains unknown, we focused on brain areas showing activation in at least one multisensory condition (p-voxel < 0.001, uncorrected). A threshold of p-voxel < 0.001, FWE corrected p-cluster < 0.05 was used to identify significant clusters. These areas were then used as regions-of-interest (ROIs) to explore the relationship between the multisensory integration of characters with children's oral reading fluency and reading comprehension proficiency. Nuisance variables of age, sex and performance IQ were controlled statistically. Bonferroni corrected $p < 0.05$ was used to address multiple comparison errors. For regions showing a significant relationship with

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reading, we additionally investigated whether congruent or incongruent condition significantly contributed to the correlation, as well as calculated the correlation between the congruency contrast with PA, RAN and MA, to reveal the cognitive-linguistic component that was most likely involved in the integration. Finally, we conducted voxel-wise whole-brain correlation as the complementary approach. All the analysis steps were replicated for *pinyin* conditions.

Brain results were presented over a surface template (smoothed ICBM152) with BrainNet Viewer (Xia, Wang, & He, 2013). Significant clusters were reported in MNI coordinates, and corresponding brain regions were localized and reported with the AAL atlas implemented in DPABI (http://rfmri.org/dpabi). Behavioral analysis and ROI analysis were performed with SPSS Statistics version 24 (SPSS Inc., Chicago, IL, USA).

3. Results

3.1. Behavioral and in-scanner performance

Descriptive statistics of demographics, cognitive-linguistic and reading skills, and in-scanner task performances are presented in **Table 1**. Positive correlations were observed between all reading measures (p's < 0.05). Regarding the relation with cognitive-linguistic skills, *Word List Reading* was associated with *Digit RAN* ($r = -0.425$, $p = 0.019$), but not with *Phoneme Deletion* ($r = 0.182$, $p = 0.336$) or *Morphological Production* ($r = 0.341$, $p = 0.065$). In contrast, *Silent Reading Comprehension* was correlated with morphological production ($r = 0.542$, $p = 0.002$),

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

but not with *Phoneme Deletion* (r = -0.060, p = 0.754) or *Digit RAN* (r = -0.077, p = 0.686). This pattern indicates that in typically developing Chinese children who have received formal reading instruction for approximately 4 years, fluent oral reading necessitates print-to-sound mapping, while proficient reading comprehension relies more heavily on semantic processing.

Regarding the in-scanner cover task, all children performed well (ACC: *M* [*SD*] = 96.7% [2.8], RT: *M* [*SD*] = 505 [68] ms). There were no significant differences in ACC (F = 0.29, p = 0.836) or RT (F = 2.59, p = 0.058) across runs, indicating that participants focused their attention on incoming stimuli during the entire scan session.

3.2. Brain results: character

In the voxel-wise whole-brain analysis, a significantly reverse congruency effect (avC < avI) was found in left IFG (peak MNI -48 16 18) and bilateral STC (peak MNI, left: -54 32 0, right: 58 -20 0; **Figure 2A and B, Table 2**) at a FWE corrected threshold of p-cluster < 0.05. We defined these clusters as ROIs to further explore their relationship with reading and found that the congruency contrast in the left IFG was positively associated with reading comprehension proficiency (Bonferroni corrected $p < 0.05$; **Figure 2C and D**), i.e., the greater the avC > avI, the better the reading comprehension. The same positive correlation was also found between congruency contrast and MA (Bonferroni corrected $p < 0.05$; **Figure 2E**), i.e., the greater the avC > avI, the better the MA. Interestingly, while the congruency-

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

comprehension correlation was driven by individual differences in activation in the incongruent condition (avC: $r = 0.164$, $p = 0.413$; avI: $r = -0.547$, $p = 0.003$), no significant correlation was found between MA and activation in either avC ($r = 0.326$, $p = 0.097$) or avI condition ($r = -0.210$, $p = 0.292$). Finally, complementary voxel-wise whole-brain correlation analysis revealed that reading comprehension positively correlated with congruency contrast in the left IFG and inferior parietal lobule (IPL) at a FWE corrected threshold of p-cluster < 0.05 (**Figure 3**, **Table 2**). In line with the above analyses, such relationships were driven by the negative correlation between reading and activation in the incongruent condition. Moreover, congruency-MA correlations in both regions were significant (left IFG: $r = 0.593$, $p = 0.001$; left IPL: $r = 0.573$, $p = 0.002$). No additional regions were identified in whole-brain regression for oral reading fluency.

3.3. Brain results: ***pinyin***

Unexpectedly, no regions showed a significant congruency effect when the whole brain analysis was performed across the entire group. However, voxel-wise correlation analysis revealed that oral reading fluency positively correlated with the congruency contrasts (avC-avI) in left STG (peak MNI -56 -36 14) (FWE corrected p-cluster < 0.05; **Figure 4A and B**, **Table 2**). ROI analysis revealed that this relationship was driven more by the avC condition ($r = 0.399$, $p = 0.039$), but not the avI condition ($r = -0.277$, $p = 0.162$). No significant correlations were found between congruency contrasts and cognitive-linguistic skills. Finally, complementary analyses was performed by dividing participants into three groups based on their

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

oral word fluency (good [n = 8]: standardized score > 1, average [n = 10]: 0 < standardized score < 1, and poor [n = 12]: standardized score < 0). Positive congruency (avC > avI) were found in all eight children in the good group ($t = 5.753$, $p < 0.001$; **Figure 4C**, **Table S1**), while negative patterns (avC < avI) were observed in 10/12 children with poor reading fluency ($t = -3.635$, $p = 0.004$). No significant effect existed in the average group ($t = 0.345$, $p = 0.738$).

4. Discussion

Establishing efficient print-to-sound mapping is essential for successful reading acquisition (Blomert, 2011; Richlan, 2019). In alphabetic languages, the congruency effect during audiovisual integration has repeatedly been observed in STC, while the direction of effect is influenced by orthographic depth. Against the background that the vast majority of research is restricted to alphabetic scripts, the present study investigated the neurofunctional basis underlying audiovisual integration of characters (deep orthography) and *pinyin* (transparent alphabetic coding system) in typically developing Chinese children for the first time. The results in both scripts are in line with predictions according to the orthographic depth hypothesis, supporting a universal principle underlying the neural mechanism of audiovisual integration of print and speech sound across languages. Moreover, we ascertained that left IFG links multisensory integration of characters and reading comprehension proficiency through automatic semantic access, while left STG links

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

multisensory integration of *pinyin* and oral reading fluency possibly through a shared common mechanism of grapho-phonological mapping.

4.1. Left IFG and bilateral STC are involved in implicit integration of multisensory information about Chinese characters

To date, various neuroimaging tools have been used to explore neural mechanisms underlying audiovisual integration of print and speech sound. Techniques such as electroencephalogram (EEG) and MEG have been used to focus on the temporal dimension (Froyen, van Atteveldt, & Blomert, 2010; Froyen, Van Atteveldt, Bonte, & Blomert, 2008; Froyen, Bonte, van Atteveldt, & Blomert, 2009; Raij, Uutela, & Hari, 2000; Xu et al., 2019). These studies demonstrated that such integration is preconscious in expert reading, which is reflected in a component of mismatch negativity/response. On the other hand, fMRI research provides greater information about spatial localization. These studies repeatedly observed difference in brain responses to audiovisual congruent versus incongruent stimulus pairs in STC that mainly consists of superior temporal gyrus and sulcus (van Atteveldt et al., 2004; van Atteveldt, Formisano, Blomert, & Goebel, 2007). Importantly, such a congruency effect has been confirmed in passive and irrelevant paradigms, pinpointing the automatic nature of this effect (Blau et al., 2010; van Atteveldt et al., 2007). Furthermore, this effect is modulated by linguistic features (Holloway et al., 2015), and is associated with individual characteristics such as reading (Blau et al., 2010; Blau et al., 2009; Karipidis et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consistency of mapping between grapheme and phoneme is an important feature in alphabetic languages. Differences in the directionality of the effect and brain regions that show this effect have been found between languages with varying orthographic depth (Blomert & Froyen, 2010; Holloway et al., 2015). In particular, greater activation in congruent condition compared with incongruent condition is more consistently reported in transparent scripts, while the opposite pattern is observed in opaque scripts. Chinese is a logographic language with an extremely deep orthography. Therefore, a reverse congruency effect can be expected based on the orthographic depth hypothesis. In support of this idea, the present study revealed three regions that showed a significantly reverse congruency effect, including STC in both hemispheres, which have been repeatedly reported in previous audiovisual integration studies. Moreover, deep orthography is not the only linguistic feature that matters. Chinese is a morpheme-based language in which semantic information is deeply involved, even at the basic character level. That is, in addition to grapho-phonological mapping, semantic access is an equally important aspect of reading (Bi, Han, Weekes, & Shu, 2007; Dang, Zhang, Wang, & Yang, 2018; Liu et al., 2017; Ruan et al., 2018; Yang et al., 2012). In the current study, the left IFG displayed a significantly incongruency effect, suggesting that this region is associated with multisensory integration. More importantly, such an effect is strongly associated with reading comprehension proficiency as well as morphological processing. In previous research, frontal region has also been reported, but along with a broad bilateral fronto-parietal network. Its functional

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

role was proposed as task related top-down modulation (van Atteveldt et al., 2007) or implicit domain-general conflict detection and resolution (Holloway et al., 2015). Recently, a reverse congruency effect was observed in frontal cortex in experienced readers in Chinese, a non-alphabetic language, with MEG (Xu et al., 2019). Here, we not only replicate such finding in a developmental population in Chinese with fMRI, but further demonstrates the role of left IFG in the implicit audiovisual integration of character is domain-specific and the information automatically processed in this region corresponds to semantics.

In addition to left IFG, one more region displayed significant correlation between congruency contrast and reading skills. In particular, congruency contrasts in the left IPL were positively associated with reading comprehension proficiency and morphological processing but not with oral reading fluency, phonological processing or naming speed. This finding is in line with the role of this region in semantic processing (Binder & Desai, 2011; Wang, Zhao, Zevin, & Yang, 2016). The left IPL has not been identified to be associated with audiovisual integration previously, however a recent training study revealed that this region was involved in the acquisition of novel grapho-phonological mapping (Xu, Kolozsvari, Oostenveld, & Hamalainen, 2020). Based on our and previous findings, it is reasonable to suggest that the functional role of a region in audiovisual integration is influenced by the linguistic features in the given language.

Taken together, these findings indicate that typically developing children in upper elementary grades integrate high frequency characters and corresponding

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

speech sounds implicitly. Moreover, the linguistic features of Chinese writing system drive the precise neural manifestation of the congruency effect, in line with previous literature, supporting the notion that while the general neural circuit for reading is universal, linguistic features in a given language introduce different demands for specific cognitive components (Bolger, Perfetti, & Schneider, 2005; Rueckl et al., 2015).

4.2. Left STG links implicit print-to-sound integration of *pinyin* with fluent oral reading

Well-established print-to-sound mapping is essential for learning to read (Blomert, 2011; Richlan, 2019). While letter knowledge is one of the most important prerequisites in alphabetic languages, *pinyin* – a transparent phonological coding system plays a scaffolding role at the initial stage in Chinese reading acquisition. Therefore, the present study, on the one hand utilizes a unique opportunity to examine the orthographic depth hypothesis with both deep and shallow orthographies in one language. On the other hand, it provides information that helps illustrate the role of *pinyin* in developing fluent reading ability.

First, as proposed by the orthographic depth hypothesis, scripts with transparent GPC should show a stronger brain activation in congruent condition compared to incongruent condition in STC. At first glance, no significant congruency effect was observed across the entire group. However, by grouping the participants based on oral reading abilities, a clear pattern appeared in the left

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

STG. In particular, all the children in the better fluency group showed higher activation in congruent versus incongruent conditions, similar to that observed in shallow orthographies (van Atteveldt et al., 2004). Moreover, average readers showed less congruency, while almost all below-average readers displayed a negative pattern. This is in line with studies in individuals with RD in these languages, where a less congruency effect was found (Blau et al., 2010; Blau et al., 2009). In terms of the brain region, the left STG is the central area that represents phonological information (Boets et al., 2013; Glezer et al., 2016). Its activation and connection patterns are known to change when learning visual-sound mappings (Dehaene et al., 2010; Li, Xu, Luo, Zeng, & Han, 2020; Thiebaut de Schotten, Cohen, Amemiya, Braga, & Dehaene, 2014), which can further scaffold later reading development (Wang, Joanisse, & Booth, 2019). It should be noted that the current conclusion can only be made for typical readers since these below-average children performed less well in the oral word reading task but were not classified as having RD.

In addition to supporting the orthographic depth hypothesis, the significant correlation between congruency contrasts in the left STG to audiovisual stimulus pairs of *pinyin* with oral reading fluency may imply that learning *pinyin* help typical children develop fluent oral reading ability (though causal claims cannot be made from this study). In comparison to other cognitive-linguistic skills, such as PA and RAN, less attention has been paid to the role of *pinyin* processing in Chinese reading development, probably because it is a scaffold in reading acquisition but not

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

considered reading itself. However, although the number of studies is limited, the *pinyin*-reading relationship has been demonstrated at the behavioral level (Lin et al., 2010; Pan et al., 2011; Siok & Fletcher, 2001). In a longitudinal study, Pan et al. demonstrated that invented *pinyin* spelling at age 6 independently predicted later reading performance at ages 8 and 10 (Pan et al., 2011). Several explanations are proposed based on behavioral evidences. One possibility is that *pinyin* can help to establish better phonological representations (McBride, Wang, & Cheang, 2018). Alternatively, training on *pinyin* can uniquely improve character-to-syllable mapping, revealed in research focusing on second language learner of Chinese (Guan et al., 2011). In this study, the congruency contrasts in the left STG to *pinyin* were positively correlated with oral reading fluency, while a similar region also showed multisensory integration of audiovisual pairs of characters. While conclusion cannot be drawn from lack of evidence, our results that brain measures were not correlated with auditory phonological processing (i.e., PA), may support the second hypothesis that learning *pinyin* shapes the neural circuit that will later be recruited in learning characters.

Overall, this study further supports the orthographic depth hypothesis by extending it to *pinyin*, an alphabetic coding system used for representing pronunciations of characters in Chinese. It also suggests that even though Chinese is a logographic writing system, print-speech sound integration still plays a critical role in reading development, in which the underlying grapho-phonological route

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

may be shaped by learning through *pinyin* prior to formal character reading instruction.

4.3. Caveats and future directions

While the current findings are appealing, several caveats should be noted. First, in this study, we investigated the neurofunctional basis underlying audiovisual integration of *pinyin* for the first time. However, we did not observe a significant congruency effect at the group level, as proposed by the orthographic depth hypothesis. This may be because while all the participants had typical reading skills, there were still individual differences in *pinyin* reading that we did not assess directly. Nevertheless, a congruency effect was found in the fluent readers, while a reverse pattern was observed in slow readers. Moreover, previous longitudinal studies have revealed a significant role of *pinyin* processing in reading acquisition (Lin et al., 2010; Pan et al., 2011; Siok & Fletcher, 2001). Taken together, the results of this study propose the left STG as the possible neural basis underlying such linkage, which should be examined in future studies. Second, since the participants recruited in this study were in grades 3-5, *pinyin* may play a different role in these individuals than it does in earlier stages of reading development. In this case, it is important to conduct a study following preliterate children until they achieve fluent reading in future. Finally, since we have uncovered brain responses to audiovisual information and their relationship with reading performance in typically developing children, such findings can be used as

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the baseline for investigating whether neural processes are altered in children with RD.

5. Conclusion

In the current study, we demonstrate a reverse congruency effect for the audiovisual integration of characters and a congruency effect for *pinyin* in typically developing Chinese children with fluent reading ability. This striking finding supports and extends the orthographic depth hypothesis. Furthermore, we found that the left IFG is strongly involved in the multisensory integration of characters and is correlated with reading comprehension and morphological awareness, suggesting that semantic representations were more automatically accessed in readers with higher comprehension proficiency. Finally, this study revealed that integrating multisensory information of *pinyin* in the left STG is correlated with oral reading fluency, implying that the grapho-phonological mapping introduced by *pinyin* learning may help children develop better reading skills.

Acknowledgements

The authors would like to thank all the children participating in this study, their families, as well as all the examiners. This work was supported by the Key Program of the National Social Science Foundation of China (14ZDB157), National Key Basic Research Program of China (grant number 2014CB846103), National Natural Science Foundation of China grants (grant number 31271082, 31671126,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

31611130107, 61374165, 81801782), Beijing Municipal Science & Technology Commission (grant number Z151100003915122), Fundamental Research Fund for the Central Universities, China Postdoctoral Science Foundation (grant number 2019T120062, 2018M641235), and Social Science Fund of Beijing (grant number 17YYA004). No conflict declared.

References

Bakos, S., Landerl, K., Bartling, J., Schulte-Körne, G., & Moll, K. (2017). Deficits in Letter-Speech Sound Associations but Intact Visual Conflict Processing in Dyslexia: Results from a Novel ERP-Paradigm. *Front Hum Neurosci, 11*(116).

Bi, Y., Han, Z., Weekes, B., & Shu, H. (2007). The interaction between semantic and the nonsemantic systems in reading: Evidence from Chinese. *Neuropsychologia, 45*(12), 2660-2673.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn Sci, 15*(11), 527-536.

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*(Pt 3), 868-879.

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Current Biology, 19*(6), 503-508.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *Neuroimage, 57*(3), 695-703.

Blomert, L., & Froyen, D. (2010). Multi-sensory learning and learning to read. *Int J Psychophysiol, 77*(3), 195-204.

Blomert, L., & Willems, G. (2010). Is there a causal link from a phonological awareness deficit to reading failure in children at familial risk for dyslexia? *Dyslexia, 16*(4), 300-317.

Boets, B., Op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., . . . Ghesquiere, P. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science, 342*(6163), 1251-1254.

Bolger, D. J., Perfetti, C. A., & Schneider, W. (2005). Cross-cultural effect on the brain revisited: universal structures plus writing system variation. *Human Brain Mapping, 25*(1), 92-104.

Chan, C. K. K., & Siegel, L. S. (2001). Phonological Processing in Reading Chinese among Normally Achieving and Poor Readers. *J Exp Child Psychol, 80*(1), 23-43.

Cui, Z., Xia, Z., Su, M., Shu, H., & Gong, G. (2016). Disrupted white matter connectivity underlying developmental dyslexia: A machine learning approach. *Human Brain Mapping, 37*(4), 1443-1458.

Dang, M., Zhang, R., Wang, X., & Yang, J. (2018). The Interaction Between Phonological and Semantic Processing in Reading Chinese Characters. *Front Psychol, 9*, 2748.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes, G., Jobert, A., . . . Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science, 330*(6009), 1359-1364.

Froyen, D., van Atteveldt, N., & Blomert, L. (2010). Exploring the Role of Low Level Visual Processing in Letter-Speech Sound Integration: A Visual MMN Study. *Front Integr Neurosci, 4*, 9.

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*(1), 23-28.

Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The Long Road to Automation: Neurocognitive Development of Letter-Speech Sound Processing. *J Cogn Neurosci, 21*(3), 567-580.

Glezer, L. S., Eden, G., Jiang, X., Luetje, M., Napoliello, E., Kim, J., & Riesenhuber, M. (2016). Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. *Neuroimage, 138*, 248-256.

Guan, C. Q., Liu, Y., Chan, D. H. L., Ye, F., & Perfetti, C. A. (2011). Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *Journal of Educational Psychology, 103*(3), 509-522.

Holloway, I. D., 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*(6), 1544-1553.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Karipidis, I. I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., . . . Brem, S. (2018). Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders. *Scientific Reports, 8*(1), 7121.

Kast, M., Bezzola, L., Jancke, L., & Meyer, M. (2011). Multi- and unisensory decoding of words and nonwords result in differential brain responses in dyslexic and nondyslexic adults. *Brain and Language, 119*(3), 136-148.

Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., . . . Zhang, Z. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology and Psychiatry, 52*(2), 212-220.

Li, H., Shu, H., McBride-Chang, C., Liu, H., & Peng, H. (2012). Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *Journal of Research in Reading, 35*(3), 287-307.

Li, M., Xu, Y., Luo, X., Zeng, J., & Han, Z. (2020). Linguistic experience acquisition for novel stimuli selectively activates the neural network of the visual word form area. *Neuroimage*, 116838.

Lin, D., McBride-Chang, C., Shu, H., Zhang, Y., Li, H., Zhang, J., . . . Levin, I. (2010). Small Wins Big:Analytic Pinyin Skills Promote Chinese Word Reading. *Psychol Sci, 21*(8), 1117-1122.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Liu, Y., Georgiou, G. K., Zhang, Y., Li, H., Liu, H., Song, S., . . . Shu, H. (2017). Contribution of cognitive and linguistic skills to word-reading accuracy and fluency in Chinese. *International Journal of Educational Research, 82*, 75-90.

McBride, C., Wang, Y., & Cheang, L. M.-L. (2018). Dyslexia in Chinese. *Current Developmental Disorders Reports, 5*(4), 217-225.

Mcnorgan, C., Randazzo-Wagner, M., & Booth, J. (2013). Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Front Hum Neurosci, 7*(388).

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Pan, J., McBride-Chang, C., Shu, H., Liu, H., Zhang, Y., & Li, H. (2011). What is in the naming? A 5-year longitudinal study of early rapid naming and phonological sensitivity in relation to subsequent reading skills in both native Chinese and English as a second language. *Journal of Educational Psychology, 103*(4), 897-908.

Plewko, J., Chyl, K., Bola, Ł., Łuniewska, M., Dębska, A., Banaszkiewicz, A., . . . Jednoróg, K. (2018). Letter and Speech Sound Association in Emerging Readers With Familial Risk of Dyslexia. *Front Hum Neurosci, 12*(393).

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron, 28*(2), 617-625.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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).

Ruan, Y., Georgiou, G. K., Song, S., Li, Y., & Shu, H. (2018). Does writing system influence the associations between phonological awareness, morphological awareness, and reading? A meta-analysis. *Journal of Educational Psychology, 110*(2), 180.

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W. J., Bick, A., Frost, S. J., . . . Frost, R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proc Natl Acad Sci U S A, 112*(50), 15510-15515.

Shah, N. J., Steinhoff, S., Mirzazade, S., Zafiris, O., Grosse-Ruyken, M. L., Jäncke, L., & Zilles, K. (2000). The Effect of Sequence Repeat Time on Auditory Cortex Stimulation During Phonetic Discrimination. *Neuroimage, 12*(1), 100-108.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child development, 74*(1), 27-47.

Shu, H., McBride-Chang, C., Wu, S., & Liu, H. Y. (2006). Understanding Chinese developmental dyslexia: Morphological awareness as a core cognitive construct. *Journal of Educational Psychology, 98*(1), 122-133.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Dev Psychol, 37*(6), 886-899.

Thiebaut de Schotten, M., Cohen, L., Amemiya, E., Braga, L. W., & Dehaene, S. (2014). Learning to read improves the structure of the arcuate fasciculus. *Cereb Cortex, 24*(4), 989-995.

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron, 43*(2), 271-282.

van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex, 17*(4), 962-974.

van Atteveldt, N. M., Formisano, E., Goebel, R., & Blomert, L. (2007). Top–down task effects overrule automatic multisensory responses to letter–sound pairs in auditory association cortex. *Neuroimage, 36*(4), 1345-1360.

Wang, J., Joanisse, M. F., & Booth, J. R. (2019). Neural representations of phonology in temporal cortex scaffold longitudinal reading gains in 5- to 7-year-old children. *Neuroimage*, *207*, 116359.

Wang, X., Zhao, R., Zevin, J. D., & Yang, J. (2016). The Neural Correlates of the Interaction between Semantic and Phonological Processing for Chinese Character Reading. *Front Psychol, 7*.

Wechsler, D. (1974). *WISC-R, Wechsler intelligence scale for children, revised*: Psychological Corporation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wu, C. Y., Ho, M. H., & Chen, S. H. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage, 63*(1), 381-391.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One, 8*(7), e68910.

Xia, Z., Hoeft, F., Zhang, L., & Shu, H. (2016). Neuroanatomical anomalies of dyslexia: Disambiguating the effects of disorder, performance, and maturation. *Neuropsychologia, 81*, 68-78.

Xia, Z., Zhang, L., Hoeft, F., Gu, B., Gong, G., & Shu, H. (2018). Neural correlates of oral word reading, silent reading comprehension, and cognitive subcomponents. *International Journal of Behavioral Development, 42*(3), 342-356.

Xu, W., Kolozsvari, O. B., Monto, S. P., & 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).

Xu, W., Kolozsvari, O. B., Oostenveld, R., & Hamalainen, J. A. (2020). Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. *bioRxiv*, 2020.2003.2012.988402.

Xu, W., Kolozsvári, O. B., Oostenveld, R., Leppänen, P. H. T., & Hämäläinen, J. A. (2019). Audiovisual Processing of Chinese Characters Elicits Suppression and Congruency Effects in MEG. *Front Hum Neurosci, 13*(18).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Xue, J., Shu, H., Li, H., Li, W., & Tian, X. J. J. o. P. R. (2013). The Stability of Literacy-Related Cognitive Contributions to Chinese Character Naming and Reading Fluency. *42*(5), 433-450.

Yan, M., Miller, K. F., Li, H., & Shu, H. (2008). What is the place for Pinyin in beginning Chinese reading? Evidence from eye movements. In *Cognitive and cultural influences on eye movements.* (pp. 343-360). Tianjin, China: Tianjin People's Publishing House.

Yang, J., Shu, H. U. A., McCandliss, B. D., & Zevin, J. D. (2012). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition, 16*(2), 354-366.

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., & Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry, 53*(8), 874-882.

Zhao, J., Wang, X., Frost, S. J., Sun, W., Fang, S.-Y., Mencl, W. E., . . . Rueckl, J. G. (2014). Neural division of labor in reading is constrained by culture: A training study of reading Chinese characters. *Cortex, 53*, 90-106.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 Demographics and behavior measures of the participants

Measures	*Mean*	*SD*	*Range*
Age (month)	127	7	111 ~ 140
Sex (female/male)	20/10		
Verbal IQ (standard score)	110	13	85 ~ 145
Performance IQ (standard score)	116	13	82 ~ 137
Full scale IQ (standard score)	114	12	93 ~ 143
Character Recognition (item)	121	11	100 ~ 146
Standard score	0.62	0.74	-0.86 ~ 2.20
Word List Reading (word / minute)	95	17	71 ~ 140
Standard score	0.49	1.14	-0.91 ~ 3.53
Reading Fluency (character / minute)	359	107	176 ~ 565
Standard score	0.93	1.08	-0.57 ~ 3.08
Phoneme Deletion (item)	21	4	7 ~ 24
Rapid Naming (second)	17	2	11 ~ 23
Morphological Production (item)	24	4	13 ~ 30

Table 2 Significant clusters identified in brain analyses

Material and contrast	Label	Brain area (AAL)	$P_{\text{FWE-corrected}}$	Size	Peak *T*	X	Y	Z
Character								
Congruency (avC vs avI)	LMTG	Left middle temporal gyrus, superior temporal gyrus	0.009	450	-5.04	-54	-32	0
	LIFG	Left inferior frontal gyrus, triangular part and pars orbitalis, precentral gyrus	0.022	303	-4.09	-48	16	18
	RSTG	Right superior temporal gyrus, middle temporal gyrus	0.043	218	-5.98	58	-20	0
Congruency (avC-avI) correlated with reading comprehension	LPrCG	Left, precentral gyrus, inferior frontal gyrus, triangular part and pars orbitalis, middle frontal gyrus	0.008	471	5.44	-32	2	36
	LIPL	Left inferior parietal gyrus, superior parietal gyrus, middle occipital gyrus, angular gyrus	0.030	307	6.53	-32	-68	46
Pinyin								
Congruency (avC-avI) correlated with oral reading	LSTG	Left superior temporal gyrus, middle temporal gyrus	0.008	397	4.88	-56	-36	14

Note: The significant clusters were identified with an FWE corrected threshold of p-cluster < 0.05 (p-voxel < 0.001). Brain area labeling is based on the AAL atlas. Cluster size refers to the number of voxels. avC = audiovisual congruent, avI = audiovisual incongruent.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1 Materials and experimental design. Example stimuli in *pinyin* (A) and character (B) conditions. (C) Schematic illustration of the fMRI procedure. (D) Auditory and visual targets in the cover task.

Figure 2 Results of brain analysis comparing between character conditions. (A) Areas showing a significant congruency effect (avC vs. avI). Threshold: p-voxel < 0.001, p-cluster < 0.05, Family-Wise Error (FWE) correction. (B) Bar plots showing activation for each condition. (C) Regions showing significant correlations with reading and related cognitive skills. (D) Scatter plot presenting a positive correlation between the congruency effect in left IFG and reading comprehension efficiency (Bonferroni corrected $p < 0.05$). (E) Scatter plot representing the positive correlation between the congruency effect in the left IFG and morphological awareness (Bonferroni corrected $p < 0.05$). Aud = auditory-only, Vis = visual-only, avC = audiovisual congruent, avI = audiovisual incongruent, L = left, IFG = inferior frontal gyrus, STC = superior temporal cortex, STG = superior temporal gyrus.

Figure 3 (A) Clusters displaying a significant correlation between the congruency contrast and reading comprehension proficiency across the brain. Threshold: p-voxel < 0.001, p-cluster < 0.05, Family-Wise Error (FWE) correction. Scatter plots displaying the correlations in left IFG (B) and the left IPL (C). IFG = inferior frontal gyrus, IPL = inferior parietal lobular.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4 (A) Clusters displaying a significant correlation between the congruency effect in *pinyin* and oral reading efficiency across the brain. Threshold: p-voxel < 0.001, p-cluster < 0.05, Family-Wise Error (FWE) correction. (B) Scatter plot displaying the correlation. (C) Bar plots showing the activation in each condition in individuals with different levels of oral reading fluency.

Figure S1 Brain activations for each condition in (A) *pinyin* and (B) character tasks. A threshold of uncorrected p-voxel < 0.001 is used for presentation purposes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

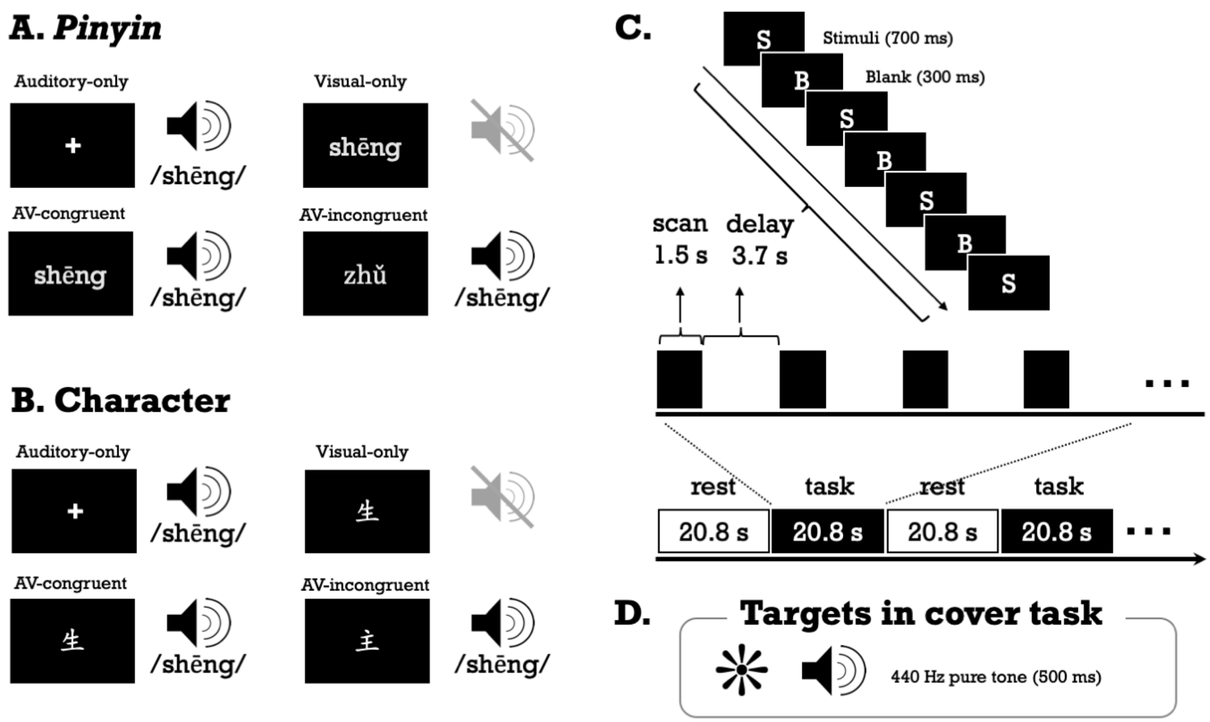

Figure 1 Materials and experimental design. Example stimuli in *pinyin* (A) and character (B) conditions. (C) Schematic illustration of the fMRI procedure. (D) Auditory and visual targets in the cover task

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

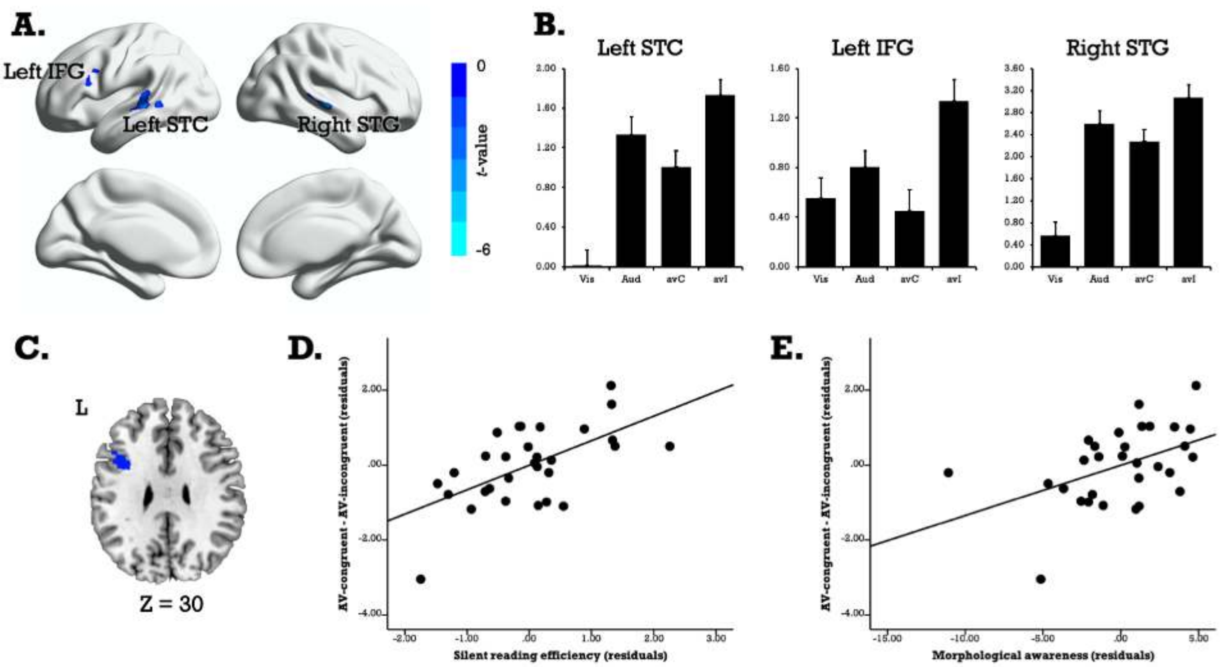

Figure 2 Results of brain analysis comparing between character conditions. (A) Areas showing a significant congruency effect (avC vs. avI). Threshold: p-voxel < 0.001, p-cluster < 0.05, Family-Wise Error (FWE) correction. (B) Bar plots showing the activation in each condition. (C) Regions showing significant correlations with reading and related cognitive skills. (D) Scatter plot presenting a positive correlation between the congruency effect in left IFG and reading comprehension efficiency (Bonferroni corrected $p < 0.05$). (E) Scatter plot representing the positive correlation between the congruency effect in the left IFG and morphological awareness (Bonferroni corrected $p < 0.05$). Aud = auditory-only, Vis = visual-only, avC = audiovisual congruent, avI = audiovisual incongruent, L = left, IFG = inferior frontal gyrus, STC = superior temporal cortex, STG = superior temporal gyrus.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

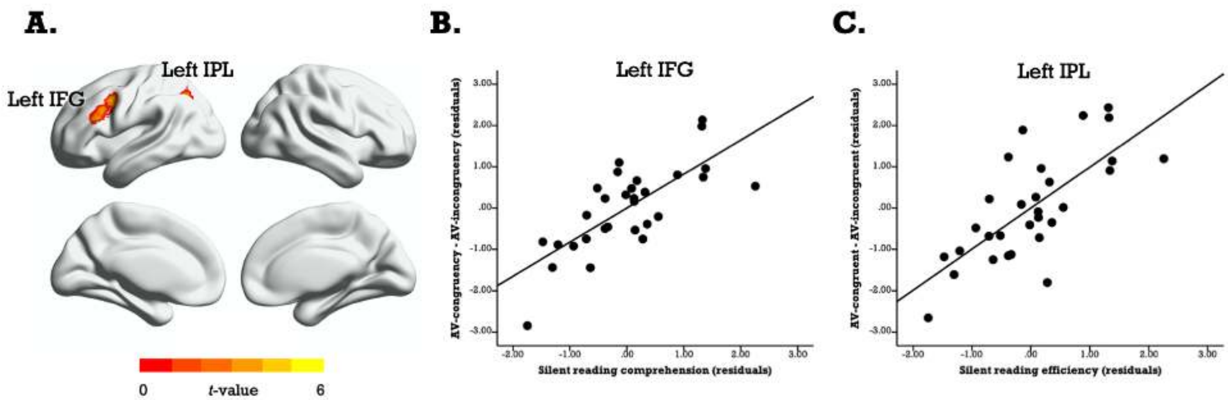

Figure 3 (A) Clusters displaying a significant correlation between the congruency contrast and reading comprehension proficiency across the brain. Threshold: *p*-voxel < 0.001, *p*-cluster < 0.05, Family-Wise Error (FWE) correction. Scatter plots displaying the correlations in the left IFG (B) and left IPL (C). IFG = inferior frontal gyrus, IPL = inferior parietal lobular.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.31.126128; this version posted June 1, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

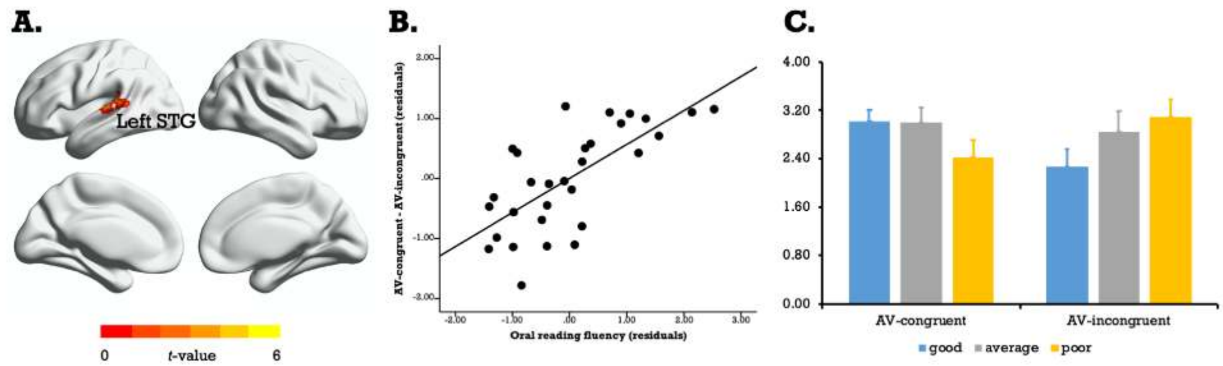

Figure 4 (A) Clusters displaying a significant correlation between the congruency effect in *pinyin* and oral reading efficiency across the brain. Threshold: p-voxel < 0.001, p-cluster < 0.05, Family-Wise Error (FWE) correction. (B) Scatter plot displaying the correlation. (C) Bar plots showing the activation in each condition in individuals with different levels of oral reading fluency.