Published in final edited form as:

Brain Lang. 2015 February ; 141: 110–123. doi:10.1016/j.bandl.2014.12.002.

Skill Dependent Audiovisual Integration in the Fusiform Induces Repetition Suppression

Chris McNorgan^{a,1} and James R. Booth^{a,b}

^aNorthwestern University

^bThe University of Texas at Austin

Abstract

Learning to read entails mapping existing phonological representations to novel orthographic representations and is thus an ideal context for investigating experience driven audiovisual integration. Because two dominant brain-based theories of reading development hinge on the sensitivity of the visual-object processing stream to phonological information, we were interested in how reading skill relates to audiovisual integration in this area. Thirty-two children between 8 and 13 years of age spanning a range of reading skill participated in a functional magnetic resonance imaging experiment. Participants completed a rhyme judgment task to word pairs presented unimodally (auditory- or visual-only) and cross-modally (auditory followed by visual). Skill-dependent sub-additive audiovisual modulation was found in left fusiform gyrus, extending into the putative visual word form area, and was correlated with behavioral orthographic priming. These results suggest learning to read promotes facilitatory audiovisual integration in the ventral visual-object processing stream and may optimize this region for orthographic processing.

Keywords

multisensory integration; reading development; fMRI; priming; orthographic processing; visual word form area

1. Introduction

We use different sensory systems to interact with different physical systems. For example, our visual system is sensitive to different wavelengths of electromagnetic energy, and our auditory system is sensitive to vibrations transmitted through a medium. Representations of these different physical systems (e.g., light, sound) thus encode very different properties.

© 2014 Elsevier Inc. All rights reserved.

Corresponding Author: Chris McNorgan, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Frances Searle Building, Room 2-328, 2240 Campus Drive, Evanston, IL 60208, Phone: 847-467-6842, Fax: 847-491-4975, chris.mcnorgan@alumni.uwo.ca.

¹Present address: Department of Psychology, University at Buffalo, The State University of New York, Park Hall Room 353, Buffalo, NY 14260-4110

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Many cognitive tasks, however, depend on multisensory integration. Thus arises the non-trivial problem of aligning different representational systems: What is a neuron that encodes line orientation to do with input from neurons that encode sound? Reading is fundamentally audiovisual, entailing mapping between visual orthographic representations and auditory phonological representations, but it emerges relatively late, and only after deliberate effort. Developmental studies of audiovisual processing during reading thus provide important insight into how experience shapes the emergence of multisensory integration. In a developmental neuroimaging study, we provide evidence that supports the argument that learning to read promotes the influence of phonological input on visual object stream to facilitate orthographic processing of visual word forms.

Orthographic processing depends critically on left fusiform gyrus (FG) (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Tsapkini & Rapp, 2010). This region demonstrates a posterior to anterior orthographic sensitivity gradient: Posterior regions are sensitive to basic orthographic features such as line conjunctions that distinguish among letters, and more anterior regions, including the putative visual word form area (VWFA) are sensitive to increasingly complex orthographic configurations such as bigrams and trigrams (Dehaene, Cohen, Sigman, & Vinckier, 2005). Because reading maps print to existing phonological representations, it follows that specialization for orthographic processing depends critically on phonological skill (Share, 1995). Importantly, despite the theoretical importance of cross-modal input in recruiting this region for reading, evidence of audiovisual processing within FG is mixed. Brem et al. (2010) found an increase in print sensitivity, marked by a greater response to words versus false fonts, in bilateral FG following training in phoneme-grapheme correspondences in pre-readers. As a likely consequence of both the participants' reading level, and the emphasis on graphemes, rather than whole-words, these differences were most pronounced in a region posterior to that typically associated with the VWFA. Nonetheless, these results are strongly supportive of a role of phonological knowledge in shaping visual processing within the FG. Evidence of phonological processing in older readers within FG has been shown only outside the VWFA (Blau, van Atteveldt, Formisano, Goebel, & Blomert, 2008; Cohen, Jobert, Le Bihan, & Dehaene, 2004; Yoncheva, Zevin, Maurer, & McCandliss, 2010), which has consequently been characterized as a "unimodal area" (Cohen et al., 2002, p. 1055). In contrast, Price and Devlin (2003) argue the VWFA is neither unimodal, nor is it specialized for visual word processing but instead falls within a multisensory convergence zone (Price & Devlin, 2011). By this account, VWFA functionality is the product of heteromodal modulation of visual-object processing.

1.1 Assessing Multisensory Integration Effects with Functional Neuroimaging

At the neural level, multisensory integration refers to a neural response function that is an interaction among two or more sensory inputs. That is, integration is assumed to have occurred when there is a significant difference in the neural response (e.g., stronger, weaker, longer, etc.) when multiple sensory inputs are present compared to when the sensory inputs are presented in isolation (Stein & Stanford, 2008). Although multisensory integration is relatively straightforward to assess using single-cell recordings, the low spatial and temporal resolution of noninvasive measures such as functional neuroimaging presents a challenge for

the design and interpretation of experiments using these approaches (Goebel & van Atteveldt, 2009). A single voxel may sample multiple exclusively unisensory sub-populations. The summed response of these sub-populations within such a voxel should be greater for multimodal stimuli. This is because multimodal stimuli are relevant to both sub-populations, whereas unimodal stimuli are relevant only to neurons encoding the corresponding modality (Laurienti, Perrault, Stanford, Wallace, & Stein, 2005). In response to this ambiguity, several alternative criteria have been proposed for using functional neuroimaging to investigate multisensory integration (Beauchamp, 2005).

The super-additivity criterion requires significantly greater crossmodal activation than the summed unimodal activations, which cannot be explained as the result of the summed activation of intermixed unisensory neurons as in the example above. According to the principle of *inverse effectiveness*: multisensory enhancement is most clearly seen where the corresponding unimodal effects are weak (Stevenson & James, 2009). In statistical terms, inverse effectiveness is a ceiling effect: Strong unimodal activation leaves little room for super-additive multisensory enhancement. Arguably as important are sub-additive multisensory interactions in which the multisensory response is reduced compared to the unimodal conditions (Noppeney, 2012). Sub-additive responses may signal response suppression, for example, following repeated activation of the same neural population from two modalities (Wiggs & Martin, 1998). According to the sharpening hypothesis (Grill-Spector, Henson, & Martin, 2006), an auditory stimulus may activate and tune a multisensory neural population encoding that item. The probability that these neurons will subsequently fire in response to the item is proportional to their roles in its encoding. A visual representation of that item will thus be less likely to activate item-irrelevant neurons, leading to a decreased response. There are, however, other non-suppressive multisensory interactions that may give rise to sub-additive responses. For example, in a facilitation model of repetition effects (Grill-Spector et al., 2006), the temporal resolution of the BOLD signal measures neural processing over a relatively wide time window. Multisensory facilitation manifesting as faster processing (e.g., priming) would result in lower net activation, calculated over the entire time window, compared to unfacilitated processing. This is because the metabolic demands for faster processes should end sooner, leading to a smaller BOLD response within that time window, though the peak neural responses for facilitated and unfacilitated conditions may be identical. Though the sharpening and facilitation models involve an overall suppression (i.e., a comparative decrease) of the BOLD signal, it is attributable to inhibitory processes only in the former. We will thus refer to these effects generally as sub-additive because our experimental design does not unambiguously distinguish between their potential neural-level causes.

Two of the most replicable effects in neuroimaging are the robust activation of left FG during decoding of printed words (Turkeltaub, Eden, Jones, & Zeffiro, 2002) and of left superior temporal gyrus (STG) during phonological decoding of speech sounds (Turkeltaub & Coslett, 2010). If super-additive effects largely depend on inverse effectiveness, multisensory integration would thus be unlikely to be associated with super-additive enhancement in either FG or STG. The literature bears this out: Neuroimaging studies of letter-speech sound integration have thus far failed to reliably demonstrate audiovisual

super-additive enhancement in left FG, even in studies where the experimental task has induced enhancement in other regions implicated in language processing (van Atteveldt, Formisano, Goebel, & Blomert, 2004). In contrast, van Atteveldt et al. (2004) found a subadditive audiovisual response within left STG, when visually-presented letters were matched with congruent (but not incongruent) speech sounds. Thus, we would predict that crossmodal modulation in either region would most likely manifest as sub-additive multisensory responses, as described by Sugihara, Diltz, Averbek, and Romanski (2006).

1.2 Does Reading Increase Multisensory Sensitivity in the Visual-object Processing Stream?

The arguments that reading experience promotes the emergence of a print-sensitive region (Dehaene et al., 2005), and that involvement of FG in orthographic processing depends on heteromodal input (Price & Devlin, 2011) hinge on the availability of evidence showing that this area is involved in multisensory processing. Though skilled reading theoretically depends on phonological knowledge (Share, 1995), to date, there is little neuroimaging evidence demonstrating FG involvement in letter-speech sound integration: Blau et al. (2008) found that FG was sensitive to audiovisual congruency when stimuli were degraded, though they did not report any multisensory modulation of overall activity, nor did they examine the VWFA in detail. Yoncheva et al. (2010) reported an attention-dependent activation of VWFA, and suppression of peripheral regions within FG during speech processing. This shows that phonological information can modulate activity along the length of the orthographic processing gradient, including, in principle, the VWFA. Nonetheless, evidence of audiovisual integration in VWFA during orthographic processing has not been convincingly demonstrated. One explanation is that, consistent with the interactive account, multisensory integration in VWFA is a product of a dynamic interaction. That is, audiovisual integration in this region might be task- and/or reader-dependent. This suggests that an exploration of multisensory integration in this region might be best viewed through the lens of a dependency on skill or performance.

In the current study, readers spanning a range of skill levels participated in a reading task in two unimodal and one cross-modal condition, while we measured brain activity with functional magnetic neuroimaging (fMRI). We examined a network of regions that previous studies of multisensory integration have shown to be involved in audiovisual integration. We predicted that the cross-modal task would elicit additive or super-additive enhancement in posterior superior temporal sulcus (pSTS), where previous research has shown additive audiovisual integration effects (e.g., Beauchamp, Lee, Argall, & Martin, 2004; Calvert, Campbell, & Brammer, 2000; Tyll et al., 2013). Conversely, if FG is involved in multisensory integration, the reliably strong recruitment of this region typically found in unimodal reading tasks suggests that the cross-modal task would elicit a sub-additive response in FG. Previous demonstrations of phonological activation of VWFA by Yoncheva et al. (2010) suggest that this sub-additive response might extend to the VWFA. If multisensory integration in this region is related to learning to read, sub-additivity in FG should be correlated with reading-skill and performance. Such a relationship would suggest that skill-dependent audiovisual integration sensitizes the visual-object processing stream for processing orthographic representations, and provides an important theoretical bridge

between early verbal development and learning to read. We similarly predicted a subadditive response in left STG because this region is associated with robust activation during unimodal auditory processing. However, because the phonological processing system develops early on, it should be relatively mature within our sample (McNorgan, Alvarez, Bhullar, Gayda, & Booth, 2011), thus the predicted subadditive response should not be skill-dependent.

2. Methods

2.1 Participants

Thirty-two right-handed children participated in the study, though four participants were excluded from the analyses because of fMRI data quality issues. The 28 retained participants were between 8.2 and 13.7 years of age ($M = 11.2$ yrs; $SD = 1.8$ yrs). As part of our participant screening process, we evaluated verbal and nonverbal intelligence quotient (IQ) using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). This test provides a brief, reliable measure of overall cognitive ability by assessing vocabulary, abstract verbal reasoning, spatial ability and abstract nonverbal reasoning. Verbal IQ ranged from 86 to 150 ($M=118.9$, $SD=15.7$) and nonverbal IQ ranged from 75 to 141 ($M=114.4$; $SD=13.6$). Scaled Letter-Word Identification sub-test scores from the Woodcock-Johnson-III Tests of Achievement (WJ-III; Woodcock et al., 2001) were used as a measure of reading skill for the purposes of participant screening. This sub-test assesses recognition of and phonological access to pronunciations associated with visual word forms, and the scaled scores ranged from 86 to 130 ($M=107.6$, $SD=12.9$). We used these scores to ensure that no participant was below 1SD from the mean for their age group with respect to his or her ability to map from print to sound. With respect to our analyses, we predicted a monotonic relationship between multisensory integration and mapping between print and speech at a particular level of fluency. Younger children reading at a particular level receive higher scaled scores than an older child reading at the same level of proficiency. Moreover, younger children with a particular scaled score read with less skill than do older children with the same scaled reading score. For the purposes of analysis, we were thus primarily interested in raw rather than scaled (i.e., peer-relative) measures of reading skill because we were interested in absolute, rather than relative, reading skill.

2.2 Experimental Task

On each lexical trial, participants were presented with paired stimuli, the order of which was counterbalanced across participants. For each scanning session, stimuli were presented in 1 of 3 modality conditions: In the cross-modal auditory/visual (AV) condition, the first item was presented auditorily, and the second was presented visually. In the unimodal auditory/auditory (AA) and visual/visual (VV) conditions, both items were presented in the auditory or visual modalities, respectively. All stimuli were presented in all three modality conditions, and thus the modality manipulation was both within-subject and within-item. Figure 1 illustrates the sequence of events within a single trial for the three modality conditions. Pairs of stimuli either rhymed or did not rhyme, and participants were asked to make a rhyme judgment response by pressing 1 of 2 keys on a handheld keypad. Participants were asked to respond as quickly and as accurately as possible, using the right index finger

for a yes (rhyme) response and right middle finger for a no (nonrhyme) response. Task selection was motivated by two primary factors: First, accurate rhyming judgments require phonological decoding of the visually-presented items, which should ideally promote audiovisual processing in the AV condition. Second, even our poorest readers would be knowledgeable of rhyming, because pre-readers have been shown to be capable of performing rhyming tasks, which correlates with subsequent single-word reading (Goswami, 1999).

Each stimulus item was presented for 800 ms, separated by a 200 ms interstimulus interval. Auditory stimuli ranged from 500 to 700 ms in duration ($M=646$ ms, $SD=64$ ms). Participants were free to respond as soon as the second stimulus item was presented. A red cross appeared for 2200 ms following the presentation of the second word, signaling to the participants to respond if they had not already done so. Responses made after the red cross disappeared from the screen were not recorded and counted as errors. A jittered response interval duration of between 2200 and 2800 ms was used to allow for deconvolution of the signal associated with each condition.

Stimulus pairs varied in terms of their orthographic and phonological similarity, and were presented in 1 of 4 similarity conditions (24 pairs per condition). There were 2 phonologically similar (i.e., rhyming) conditions, one with orthographically similar pairs (O+P+; e.g., CAGE-RAGE) and another with orthographically dissimilar pairs (O-P+; e.g., GRADE-LAID). There were also 2 phonologically dissimilar (i.e., nonrhyming) conditions, one with orthographically similar pairs (O+P-; e.g., STAMP-SWAMP) and one with orthographically dissimilar pairs (O-P-; e.g., THIEF-PLEAD). Stimulus characteristics for the four similarity conditions is presented in Table 1. The conditions were generally comparable with respect to mean word length, frequency or naming accuracy. Though written frequency tended to be higher for non-rhyming pairs than for rhyming pairs, the conditions did not differ significantly with respect to frequency ($F(3,92) = 1.37, p=.26$). Moreover, as we explain below, we restricted our analyses to the two rhyming condition, which were even more closely matched ($t(46)=-.71, p=.48$). Thus, congruency conditions were not distinguishable by word frequency. Finally, our analyses focused on differences between modality conditions. Because modality conditions used the same stimuli, they were necessarily matched on stimuli characteristics. The word stimuli used in our experiment are listed in Appendix A.

Fixation trials (24 for each run) were included as a baseline and required the participant to press the “yes” button when a fixation cross at the center of the screen turned from red to blue. Perceptual trials (12 trials for each run) were included as a localizer condition for a related study (McNorgan, Awati, Desroches, & Booth, 2013). Perceptual trials comprised 2 sequences containing tones (AA), nonalphabetic glyphs (VV), or tones followed by glyphs (AV). These stimuli were presented as increasing, decreasing or steady in pitch (for auditory stimuli) or height (for visual stimuli). Analyses were carried out on only the rhyming (i.e., “yes”) items to avoid introducing confounding influences arising from making “yes” versus “no” decisions, and because controlling phonological similarity permitted an analysis of the orthographic priming effect (see below).

2.3 MRI Data Acquisition

Participants were positioned in the MRI scanner with their head orientation secured using foam pads. An optical response box was placed in the participant's right hand to log responses. Visual stimuli were projected onto a screen, which participants viewed via a mirror attached to the inside of the head coil. Images were acquired using a 3.0-Tesla Siemens Trio scanner. The blood oxygen-level-dependent (BOLD) signal was measured using a susceptibility weighted single-shot echo planar imaging (EPI) method. Functional images were interleaved from bottom to top in a whole-brain acquisition. The following parameters were used: TE = 20 ms, flip angle = 80 degrees, matrix size = 128×120, field of view = 220×206.25 mm, slice thickness = 3 mm (0.48 mm gap), number of slices = 32, TR = 2000 ms. Before functional image acquisition, a high resolution T1-weighted 3D structural image was acquired for each subject (TR = 1570 ms, TE = 3.36 ms, matrix size = 256 × 256, field of view = 240 mm, slice thickness = 1 mm, number of slices = 160). Neuroimaging experiments involving auditory stimuli are sometimes designed around a sparse sampling sequence, in which auditory stimuli are presented only during quiet periods between image acquisitions. This helps avoid the problem of scanner noise interfering with perception of the auditory stimuli, but requires longer runs to acquire the same amount of data. Though the AA and AV modality conditions might have benefitted from a design employing sparse acquisition, our experience working with our experimental population has shown us that longer runs are especially difficult for our younger participants and those with reading difficulty. Longer runs are also more likely to induce fatigue. Finally, previous research showing audiovisual integration effects in the visual stream (Blau et al., 2008) used auditory stimuli that had been degraded with noise, suggesting that the scanner noise would not negatively impact our ability to detect an influence of phonological information on orthographic processing. For these reasons, we chose a rapid event-related design over a slower sparse sampling design. Participants wore sound attenuating headphones to minimize the effects of the ambient scanner noise. To minimize fatigue and accommodate participant availability, participants took part in the experiment over a number of scanning sessions within a 6-month period (median interval between scan sessions was 5.5 weeks). During each scanning session, participants completed the rhyming task for one modality condition (2 runs). In instances where the MRI preprocessing step had detected data quality issues for a previously acquired run for a participant (e.g., excessive head movement), up to one attempt was made to reacquire that data by adding an extra run to a subsequent session, time permitting. All participants underwent extensive training in a mock scanner prior to scanning and practiced the task outside the scanner immediately prior to each fMRI acquisition session. Thus, participants were familiar with the task and the scanning environment before each fMRI session.

2.4 fMRI Data Analysis

fMRI data were analyzed using Statistical Parametric Mapping (SPM8, <http://www.fil.ion.ac.uk/spm>). ArtRepair software (<http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html>) was used to correct for participant movement. Images were realigned in ArtRepair, which identified and used interpolated values from the 2 adjacent non-outlier scans to replace outlier volumes, defined as those with head movement exceeding 4 mm in any direction or deviations of more than 1.5% from the mean global

signal. No more than 10% of the volumes from each run and no more than 4 consecutive volumes for any individual were interpolated in this way. Slice timing was applied to minimize timing-errors between slices. Functional images were co-registered with the anatomical image, and normalized to the Montreal Neurological Institute (MNI) ICBM152 T1 template. Images were smoothed using a 2mm isotropic Gaussian kernel.

A general linear model (GLM) was calculated for each subject, separately modeling for each run the rhyming lexical conditions and the fixation baseline as conditions of interest, and the non-rhyming lexical conditions and three perceptual conditions as conditions of non-interest. First level contrasts compared lexical versus baseline activations within each task modality separately and for the mean additive [$AV > (AA + VV)/2$], super-additive [$AV > (AA+VV)$] and sub-additive [$AV < (AA+VV)$] audiovisual integration criteria (Beauchamp, 2005). Second-level multiple regression analyses assessed group-level enhancement (i.e. additive and super-additive) and sub-additive effects, including reading skill (as measured by the WJ-III Word ID subtest) as a regressor of interest, and task accuracy as a nuisance regressor indexing task difficulty for each participant. The influence of task difficulty was partialled out because the BOLD response has been shown to be sensitive to difficulty (Gould, Brown, Owen, ffytche, & Howard, 2003). Though both accuracy and reaction times have been shown to index task difficulty in fMRI studies (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003), reaction time was a factor in our correlational analyses. Therefore, additionally partialling out the variance associated with this variable was contraindicated. These second-level analyses were examined at the whole-brain level to assess overall task activation relative to baseline, and audiovisual additive/super-additive or sub-additive effects, identifying the unique contribution of reading skill. We used raw, rather than scaled, Word ID scores as regressors because they reflect absolute performance on a reading task. Our approach thus examined patterns of brain activity associated with reading at a particular level of skill. These analyses revealed those brain regions in which activity was either generally modulated by crossmodal stimuli or became so as a function of reading skill.

2.5 Region of Interest (ROI) definitions

Areas showing skill-dependent enhancement or sub-additivity were defined as ROIs, in which we additionally correlated mean activity with decision latency on the rhyming task. These ROI analyses permitted us to characterize the patterns of skill-dependent crossmodal modulation of neural activity by examining behavioral priming effects (O+P+ compared to O-P+).

To assess laterality effects in audiovisual additivity, we created bilaterally symmetrical ROIs comprising the intersection of any region showing either an additive or super-additive effect in each hemisphere and the contralateral mirror of these regions (Figure 3C). That is, if the voxel at $[x, y, z]$ was associated with an additive or super-additive effect, *and* the voxel at $[-x, y, z]$ was also associated with either an additive or super-additive effect, then we included those voxels in the right and left ROIs, respectively. If *either* voxel failed to show an additive or superadditive effect, then *both* $[x, y, z]$ and $[-x, y, z]$ were excluded from the ROIs. This permitted us to conservatively assess laterality in (approximately) homotopic

anatomical regions, by ensuring that our contrasts included only voxels demonstrating additivity at some statistical threshold (see Section 3.2.3).

3. Results

3.1 Behavioral Results

Mean accuracy and decision latencies by task modality for the two orthographic conditions used in the neuroimaging results that follow (O+P+ and O-P+) are presented in Table 2. Analyses of accuracy found a significant effect of task modality, $F(2,54) = 16.07$, $p < .001$, $\eta^2 = .37$, orthographic condition, $F(1,27) = 1.38$, $p = .02$, $\eta^2 = .19$, and that task modality and orthographic condition interacted, $F(2,54) = 8.51$, $p = .001$, $\eta^2 = .24$. Bonferroni post-hoc pairwise comparisons of the three task modalities indicate that the VV condition ($M = .92$; 95% CI [.89, .95]) was significantly more accurate than both the AA condition ($M = .82$; 95% CI [.77, .87]), $p < .002$ and the AV condition ($M = .81$; 95% CI [.77, .85]) $p < .001$, however the AV and AA conditions did not differ $p = 1.00$. This pattern may be attributable to the fact that the auditory presentations were often accompanied by scanner noise, making them more difficult to discriminate. Posthoc comparisons between the orthographic conditions showed participants were somewhat less accurate for the orthographically overlapping rhyming condition ($M = .84$; 95% CI [.81, .87]) than the non-overlapping condition ($M = .86$; 95% CI [.83, .89]), $p < .02$. Finally, in light of the significant interaction, post-hoc pairwise comparisons were performed between orthographic conditions for each task modality. To correct for multiple comparisons, the significance threshold for these tests was set to 0.017. Orthographic similarity conditions were equally accurate for the AV ($t(27) = -.91$, $p = .37$) and VV ($t(27) = 2.09$, $p = .05$) task modalities, whereas the O+P- condition was significantly more accurate than the O+P+ condition for the AA task modality ($t(27) = -4.47$, $p < .001$). Because this effect was counterintuitive, driven by the auditory modality, and was found among high accuracy rates, we do not feel that this behavioral effect reflects a reliable reading-related process.

Analyses of decision latencies found a significant effect of task modality, $F(2,54) = 25.14$, $p < .001$, $\eta^2 = .50$, and task modality and orthographic condition interacted, $F(2,54) = 6.89$, $p = .002$, $\eta^2 = .20$, though decision latencies did not significantly differ overall between the two orthographic conditions, $F(1,27) = .53$, $p = .47$. Bonferroni post-hoc pairwise comparisons of the three task modalities indicate that participants were slower to respond in the AA condition ($M = 1397$ ms; 95% CI [1281, 1514]) than the AV condition ($M = 1094$ ms; 95% CI [985, 1202]), $p < .001$ and the VV condition ($M = 1226$ ms; 95% CI [1120, 1331]), $p < .001$. Following the significant interaction, post-hoc pairwise comparisons were performed between orthographic conditions for each task modality. To correct for multiple comparisons, the significance threshold for these tests was set to 0.017. Overall, decision latencies were equivalent for the two orthographic conditions in the AV task modality ($t(27) = -.99$, $p = .33$), but participants were slower to respond to the orthographically overlapping condition in the AA task modality ($t(27) = 2.72$, $p = .01$) but faster to respond to the orthographically overlapping condition in the VV task modality ($t(27) = -2.60$, $p = .015$). The pattern of decision latencies suggest that rhyming judgments were facilitated by orthographic priming in the VV task modality. Because the orthographically overlapping

AA condition had both the lowest accuracy and slowest decision latency, this suggests that participants were not making a speed-accuracy tradeoff. Rather, this pattern is consistent with both increased processing effort in the presence of scanner noise and the inability to make rhyming judgments on auditory-only items before vocalization completes.

3.2 Neuroimaging Results

3.2.1 Overall Activation Within Modality—Whole-brain voxel-wise contrasts between lexical and fixation baseline trials, partialled out the effects of reading skill and task performance. [Lexical > Fixation] (*task*) contrast maps were generated separately for the AV, AA and VV modality conditions. These maps were thresholded at $p < .001$ (uncorrected) with a cluster-extent false discovery rate (FDR) correction of $q < .01$. We used the same significance threshold to calculate reverse contrast maps (i.e., [Fixation > Lexical], *baseline*), inclusively masked by the voxels included in the union of the task contrast maps for all three task modalities. This permitted detection of voxels that were activated in at least one modality and deactivated in at least one other modality. Note that by masking the baseline contrast in this way we eliminated brain regions showing universal task-related deactivations (e.g., the default mode network). The baseline contrast was carried out to facilitate the discussion of additive or subadditive effects in terms of deactivation. The associated maps are presented in Figure 2. The coordinates for the peak cluster activations for these maps appear in Table 3. The AV condition was associated with bilateral occipital visual and STG activation, left FG and ventral, middle and dorsal IFG activation. The AA condition was associated with bilateral STG activation and ventral IFG activation. Finally, the VV condition was associated with broad bilateral occipital activation, and left FG, posterior STG and ventral, middle and dorsal IFG activation. These patterns are comparable to those presented in (McNorgan, Awati, et al., 2013), which examined crossmodal influences in the auditory processing system.

3.2.2 Super-additive Audiovisual Enhancement—Whole-brain voxel-wise regression analyses, partialling out the effects of reading skill and task performance, revealed super-additive audiovisual enhancement in right inferior parietal lobule (IPL), precentral gyrus (PrG), right caudate, and a large swath of right temporoparietal cortex that included supramarginal gyrus (SMAR) but extended ventrally to pSTS and MTG (Table 4, Figure 3A). Super-additive audiovisual enhancement was not modulated by reading skill in any region, suggesting that reading experience did not lead to an increased hemodynamic response for cross-modally presented whole word stimuli.

3.2.3 Additive Audiovisual Enhancement—Because it has been suggested that the stringency of the super-additive criterion might adversely affect statistical power in neuroimaging studies of audiovisual processing (Beauchamp, 2005), we additionally assessed additive audiovisual enhancement using a hybrid of the mean and max additivity enhancement criteria. Briefly, the mean additive criterion [$AV > (AA + VV)/2$] requires cross-modal activity to be significantly greater than the average of the two unimodal conditions. This criterion could be met, however, in a unimodal region. For example, the cross-modal condition might activate unimodal visual areas to a modest degree, the visual-only condition could activate these areas robustly, and the auditory-only condition could fail

to activate them at all. We carried out a whole-brain voxel-wise regression analysis of audiovisual enhancement meeting the mean additive criterion, again partialling out the effects of reading skill and task performance. The mean additivity criterion was made more stringent by further inclusively masking voxels meeting this criterion with the intersection of regions for which the cross-modal task had numerically - but not necessarily significantly - greater activation relative to both unimodal conditions (the max additive criterion, $[AV > \text{Max}(AA, VV)]$). This mask thus ensured that all voxels demonstrating significant cross-modal enhancement effects did so with respect to both unimodal conditions, at least to some extent.

Naturally, right hemisphere regions meeting the super-additivity criterion also met the less stringent additivity criterion. Table 4 and Figure 3B additionally show left posterior middle temporal gyrus (MTG, extending to pSTS), superior parietal lobule (SPL) and inferior frontal gyrus (IFG) showed an additive cross modal response, suggesting that these regions are sensitive to both modalities and show an overall enhancement in the hemodynamic response during audiovisual integration. The additive response was not modulated by reading skill in any region.

3.2.4 Laterality Effects in Additivity and Super-additivity—As discussed previously, detecting super-additive effects often depends critically on inverse-effectiveness of the unimodal stimuli. Thus, the observation of super-additive effects in the right, but not left hemisphere may stem from the fact that language is generally left-lateralized, even in non-right-handed individuals who show a greater incidence of atypical language lateralization (Szaflarski et al., 2002; but see also Vigneau et al., 2011, for meta-analytic evidence that right inferior frontal gyrus (Pars Triangularis) contributes towards contextual processing during sentence comprehension). That is, a language task is likely to recruit left hemisphere for all task modalities, but may recruit right hemisphere only under certain conditions. This would decrease the likelihood of seeing super-additive effects in left but not right hemisphere. To assess this possibility, we used bilaterally symmetrical ROIs containing mirrored voxel pairs where both voxels exhibited additive or super-additive effects. We calculated the mean signal for the rhyming and fixation conditions within each of these ROIs for each of the conditions (Figure 4). A repeated measures analysis of variance of the cross-modal and summed unimodal signal within left and right hemispheres found a main effect of hemisphere, $F(1,27) = 93.28, p < .001$ and of modality, $F(1,27) = 18.76, p < .001$, and a significant interaction between hemisphere and modality, $F(1,27) = 33.36, p < .001$, consistent with a language-laterality related inverse effectiveness of the unimodal conditions. Within the right hemisphere, the AV condition was close to baseline, whereas the unimodal conditions show decreased activation relative to baseline. Cross-modal and summed unimodal signal significantly differed in this region, $t(27) = 6.66, p < .001$, creating super-additive cross-modal enhancement in the right hemisphere. However, this super-additive enhancement reflects a deactivation in the unimodal conditions, relative to baseline, rather than an increased activation for the cross-modal condition. In the left hemisphere, all task modalities were associated with increased activation relative to baseline, and the cross-modal and summed unimodal signals did not significantly differ,

$t(27) = 1.45, p = .32$. Thus the AV enhancement failed to reach the super-additivity criterion in the left hemisphere, though it met the less strict max additivity criterion in this region.

3.2.5 Skill-Dependent Sub-Additivity—Whole-brain voxel-wise regression analyses, partialling out the effects of reading skill and task performance failed to find a main effect of audiovisual sub-additivity in any region. Reading skill (raw Word ID scores) did, however, independently predict audiovisual sub-additivity at the whole brain level in left FG, at the junction of the posterior MTG and middle occipital gyrus, and right frontal pole (Table 5, Figure 5). The left hemisphere sub-additivity was notable because it was found along the length of the ventral object processing stream in a region supporting orthographic processing (FG) and extending into the VWFA (Figure 3A), the location of which we determined from coordinates reported in previous studies (Cohen et al., 2002; Price & Devlin, 2003). This suggests that learning to read induces a change in the top-down influence of audiovisual input into the visual system responsible for processing printed word forms. Though scaled reading scores represent a child's peer-relative reading ability by factoring out a child's age, we recalculated the whole-brain regression model using scaled scores. Critically, we again identified the same significant cluster in the left FG, overlapping with the VWFA. A significant cluster was also found right frontal pole, though the MOG/MTG cluster failed to reach significance.

To present a clearer picture of the relationship between activity and reading skill in the orthographic processing system, we measured mean signal across all lexical trials in each of the task modalities in voxels in MOG and FG that showed a correlation between sub-additivity and skill at the group level. In both of these regions, audiovisual activity is above that for both unimodal conditions in poorer readers, but drops below that for the unimodal conditions in better readers (Figure 6). We note, however, that because these regions were selected on the basis of a significant group-level correlation between reading skill and activity, statistics calculated on these values would be biased, and are thus not presented. Verbal and nonverbal IQ scores were significantly correlated among our participants ($r(26) = .54, p = .003$), though the reading skill measure was not significantly correlated with either verbal IQ ($r(26) = .26, p = .18$) or nonverbal IQ ($r(26) = .33, p = .09$). Nonetheless, we repeated the analysis of the sub-additivity in MOG and FG to rule out the possibility that these effects might be accounted for by IQ rather than reading skill. Neither region showed a significant correlation between sub-additivity and IQ (MOG: verbal IQ $r(26) = -.10, p = .61$; nonverbal IQ $r(26) = -.15, p = .44$; FG verbal IQ $r(26) = -.16, p = .42$; nonverbal IQ $r(26) = -.10, p = .63$). Thus, sub-additivity in the ventral visual processing stream appears to depend on reading skill rather than general IQ.

A cognitive interpretation of sub-additivity is that it reflects a priming effect induced by repeated presentation of representationally similar stimuli (Buckner et al., 1998; Thompson-Schill, D'Esposito, & Kan, 1999). If so, one would expect the behavioral priming effect (i.e., faster decision latencies for orthographically similar versus dissimilar pairs) in the cross-modal task to correlate with sub-additivity. We calculated the behavioral priming effect as the difference between the mean decision latency in the cross-modal task for rhyming trials with matching (O+P+) versus mismatching orthography (O-P+). Because the first item was presented auditorially in the cross-modal task, this priming effect reflects faster processing

when phonological input into the orthographic system activates matching versus mismatching orthographic representations. Pearson correlations between the behavioral priming effect for correct responses only and sub-additivity were significant in both MOG, $r(26)=.38, p=.024$ and FG, $r(26)=.32, p=.049$. Participants who showed the greatest audiovisual sub-additivity in the ventral visual-object processing stream were most sensitive to orthographic similarity in the cross-modal condition (Figure 7). Audiovisual additive (or super-additive) enhancement was not correlated with decision latency in any region. A similar analysis of task accuracy was not performed because this measure was used as an index of task difficulty, and partialled out as a nuisance variable.

4. Discussion

A rhyming judgment task for item pairs presented in two unimodal (auditory- and visual-only) and one cross-modal (audiovisual) condition revealed that audiovisual processing modulated the BOLD response in a network of brain areas implicated in skilled reading and audiovisual integration. The pattern of these results provide insight into how orthographic processing and mapping a printed word to its phonological referent results from a skill-dependent interaction between auditory and visual processing systems.

4.1 Skill-Dependent Sub-additivity

Our primary goal was to find evidence of skill-dependent audiovisual integration in the nominally unimodal orthographic processing stream during letter-speech sound processing. Our interest follows from the assumption that reading development builds upon existing phonological knowledge and yet the neuroimaging evidence is weak on this point. Our main finding was that the neural correlates of audiovisual integration manifested as a decrease in the BOLD response in the visual processing stream, including the VWFA, and that this was related to both reading skill and orthographic priming. Sub-additive interactions with equivalent responses for auditory, visual, and audiovisual conditions are most consistent with amodal processing (Noppeney, 2012). In the task used in the current study, however, whereas fusiform gyrus is very strongly active in the visual condition, it is not significantly active above baseline during the auditory condition (McNorgan, Awati, et al., 2013). This would suggest that sub-additivity in this region reflects an auditory modulation of visual processing. This finding is of particular importance because it suggests that learning to read changes processing along the length of the visual-object processing stream by incorporating phonological information, facilitating the recognition of print as language-relevant objects.

Learning to read fundamentally entails linking established phonological representations with novel visual word representations (Share, 1995). Though several neurobiological models of reading development have been proposed, all highlight the importance of the phonological system in the development of the orthographic system. Pugh et al. (2010), for example, argue that the orthographic tuning of the ventral language processing stream, which includes the FG, depends on the dorsal phonologically-tuned language system. Similarly, Schlaggar and McCandliss (2007) endorse an interactive specialization framework, in which functional specialization (in this case, of the visual processing stream for orthographic processing) is driven by interactions between co-activated brain regions. In this framework, the co-activation of the phonological and the visual systems (e.g., retrieval of candidate

phonological representations for a printed word) change existing circuits involved in visual and phonological processing and form novel functional connections among these circuits. Our results further underscore the importance of audiovisual integration in reading development, showing that integration processes are tied to activity in the ventral visual-object processing stream, and are reflected in reading skill. Indeed, it has been suggested that skill-dependent top-down influences associated with processing visual words are responsible for the pattern of decreasing activity within this region as young readers gain reading experience (Schlaggar & McCandliss, 2007). Our results suggest that this top-down influence entails audiovisual integration. Furthermore, the fact that sub-additivity is inversely related to decision latency (i.e., greater sub-additivity is associated with faster decision latencies) indicates that it is a facilitatory effect. These results additionally provide insight into our previous work, allowing us to interpret our longitudinal work examining reading development in children in the context of audiovisual integration. In McNorgan et al. (2011), an increase in FG activity for older readers was associated with smaller gains in reading skill. We interpreted increased activation in the orthographic system as reflecting additional processing associated with pattern separation for orthographic representations. Better readers, we reasoned, would more efficiently apply resources involved in phonological and audiovisual integration to assist in pattern separation, thereby decreasing the load on the orthographic system. The present results are consistent with this interpretation.

The finding that the sub-additive response in FG was correlated with behavioral orthographic priming effects, allows us to interpret FG sub-additivity as a neural priming effect. This further strengthens our argument that reading skill is linked to audiovisual integration in the visual object processing stream. Neural priming, however, may reflect any number of physical processes (Grill-Spector et al., 2006). In their review, Grill-Spector and colleagues describe three models of neural priming, each reflecting a different physical process, that account for reduced activation for repeated stimuli. The correlation between cross-modal neural and behavioral priming effects seems most compatible with two of these models. One interpretation of our results, rooted in the sharpening model of neural priming, is that skill-related sub-additivity reflects inhibition of competing orthographic representations of various grain sizes. Higher reading skill is associated with greater top-down influence of existing phonological knowledge into the orthographic processing stream. This sharpens orthographic representations in more skilled readers, decreasing ambiguity associated with competing responses and thereby enhancing performance. A second interpretation, rooted in the facilitation model, is that phonological information admitted into the orthographic processing stream in those with higher reading skill increases sensitivity of the region to congruent orthographic input. In this model, the increased sensitivity is reflected by faster firing and consequent return to baseline and thus a lower BOLD signal measured over the acquisition time window. van Atteveldt, Peterson, and Schroeder (2014) used conceptually related object sounds to produce audiovisual suppression in early visual cortex when response speed was emphasized. Los and Van der Burg (2013) used task-irrelevant tones to facilitate spatial judgments to visually-presented cues. van Atteveldt et al. describe their results in terms of audiovisual integration, whereas Los and Van der Burg argue that the facilitation they observed is the result of temporal preparation – the auditory

cue signals that a visual stimulus is imminent. This preparedness could plausibly manifest as a reduced firing threshold, reflected in a sub-additive AV response. However, as Los and Van der Burg remark, the literature presents evidence for both multisensory integration and for temporal preparation, both of which are consistent with the faster decision latencies and sub-additive BOLD response in the visual object processing stream. Importantly, both the facilitation and the sharpening interpretations predict a correlation between behavioral and neuronal priming effects. Differentiating between them rests on determining whether phonological input inhibits inconsistent orthographic representations, or excites consistent representations, though both influences may be possible, depending on stimulus or participant characteristics. Our experiment was not ideally designed to test these alternatives, though the [Fixation > Lexical] contrasts may provide some insight into the nature of the sub-additive response. In particular, we found that, among voxels showing activation in any task, only the AA task was associated with task-related deactivation, restricted to lateral and posterior regions of the middle occipital gyrus. Thus, whereas the terms ‘inhibition’ and ‘suppression’ may connote a task-related deactivation (relative to baseline), skill-related audiovisual sub-additivity in FG may be better described as a signal that becomes closer to baseline. This pattern is more consistent with the facilitation model, in which the AV condition is associated with a faster return to baseline, rather than with suppression below baseline.

Previous findings of audiovisual subadditivity in STG (van Atteveldt et al., 2004), a region implicated in phonological processing, lead us to predict a sub-additive response in this region, consistent with the principle of inverse effectiveness. Moreover, we have previously used a rhyming judgment task to show skill-dependent sensitivity to multisensory congruency in phonological processing regions (McNorgan, Awati, et al., 2013). Our present analysis found no evidence of overall crossmodal modulation of activity in either direction within this region. This prediction was not directly related to the question of the development of the orthographic processing system and therefore of secondary interest. Nonetheless, we wish to reconcile our present result with these previous findings. Importantly, van Atteveldt and colleagues examined synchronous letter/speech-sound pairs that were either congruous (e.g., ‘a’, /æ/) or incongruous (e.g., ‘a’, /ɪ/), finding evidence of multisensory integration only for congruent pairs. In contrast, the rhyming task in our study requires participants to maintain and phonologically compare two phonologically complex stimuli over more than 1000 ms. It is probable that the rhyming task places a prolonged demand on the phonological processing compared to the passive perception task used by van Atteveldt and colleagues. Additionally, statistical power considerations required us to collapse orthographic congruency conditions (O+P+ and O-P+), obscuring potential additive or subadditive modulation associated with one congruency condition but not the other. When orthographic congruency conditions are contrasted against one another (rather than individually against the unimodal tasks), we found a skill-dependent audiovisual congruency effect in the STG (McNorgan, Awati, et al., 2013). Because congruency effects are a function of the relationship between responses to congruent and incongruent stimuli *within* a modality, whereas additivity/subadditivity is a function of relationships *between* modalities, congruency and additivity effects may vary independently. Skill-dependent congruency effect for AV pairs but not for AA or VV pairs are thus consistent with super-

additive, sub-additive and null effects that may or may not correlate with skill in either direction.

Another possible explanation for the failure to find the predicted cross-modal modulation in STG is that this region is sensitive to detecting speech in noise. We interpreted the lower accuracy on auditory conditions as reflective of the challenge that the scanner noise presented to hearing the verbal stimuli. Previous research has shown that detecting speech in noise is associated with an increase in activation in STG (Wong, Uppunda, Parrish, & Dhar, 2008). Thus, identifying the auditorily-presented stimuli from in the AA and AV conditions may have lead to an increase in STG activation for both conditions. That is, speech-in-noise related activation would have made the AV and AA conditions more similar in these regions, thereby muting potential audiovisual modulation effects. Detecting speech in noise has also been associated with increases in activation in insula (Wong et al., 2008), anterior superior temporal sulcus and middle temporal gyrus (MTG) (Bishop & Miller, 2008) and left premotor cortex and dorsolateral prefrontal cortex (Scott, Rosen, Wickham, & Wise, 2004), of which only the MTG displayed audiovisual modulation in our experiment.

4.2 Additive and Super-additive Enhancement

Many measures of integration effects have been proposed (Beauchamp, 2005), and the mean criterion is the most lax of these. More stringent criteria (e.g., $[AV > \max(AA, VV)]$) require more robust integration effects for inclusion, and would certainly be more appropriate for analyses involving contrasts between integration effects (e.g., does integration occur for condition A but not for B?). However, our correlational approach asked whether reading skill is related to audiovisual integration effects, to the extent that they appear. Thus, eliminating the voxels (and possibly participants) that show the weakest integration effects works at cross-purposes to this goal. By combining elements of the mean and max additivity criteria, our hybrid additivity criterion was designed to be sensitive to weaker integration effects, for the purposes of a correlational approach, while also ensuring that the additive effects numerically (if not statistically) also met the max criterion. Concerns associated with the use of a more lax additivity criterion are mitigated somewhat by the fact that spurious estimates of additivity would, by assumption, be noise, and theoretically uncorrelated with reading measures.

Audiovisual integration has been shown to be sensitive to temporal dynamics (Stevenson & Wallace, 2013; van Atteveldt, Formisano, Blomert, & Goebel, 2007) and depend on the efficacy of the unimodal stimuli (Stevenson & James, 2009). Because the enhancement effects we found were not correlated with performance, but should depend on temporal coincidence, one explanation is that enhancement in prefrontal cortex (i.e. inferior frontal gyrus) in our study either took place after task-essential integration took place, or was ongoing during the cross-modal condition. Both of these possibilities would be consistent with the proposed role of prefrontal cortex in coordinating input from multiple modalities by mediating memory and attention (Klemen & Chambers, 2012). However, because prefrontal cortex has been shown to demonstrate very early sensitivity to audiovisual congruency (Bernstein, Auer Jr, Wagner, & Ponton, 2008), this would be most consistent with the assumption that audiovisual enhancement in prefrontal cortex was ongoing.

The laterality analysis of additivity enhancement suggests that the super-additive audiovisual enhancement observed in right (but not left) hemisphere reflects the principle of inverse effectiveness. It is well known that there is a left hemisphere dominance for language (Knecht et al., 2000), and our task required orthographic to phonologic mapping in at least one unimodal condition (VV). This would predict a stronger summed unimodal response in the left hemisphere among populations involved in orthographic to phonological mapping. Notably, the unimodal response was above baseline in the left pSTS, and marginally below baseline in right pSTS. The importance of left pSTS in language-related audiovisual integration has been well-documented (e.g., see van Atteveldt, Roebroeck, & Goebel, 2009 for a review), and the additivity effects we observed in this region are consistent with this literature. There was neither strong activation nor deactivation in the right pSTS, relative to baseline, for the cross-modal condition. Rather, this region showed some deactivation for both unimodal conditions, which summed to a negative value and appeared as a super-additive enhancement. Though audiovisual enhancement in right pSTS cannot be ruled out, it is also possible that this pattern reflects a meaningful deactivation of this region in the unimodal conditions. Disambiguating between these two possibilities, and determining what role, if any, unimodal deactivation of the right pSTS might play in language processing is beyond the scope of the present study. That clear laterality differences emerged in pSTS, however, suggests that unimodal and cross-modal processing proceeds differently in right and left pSTS, and suggests this is a finding that should be further explored.

4.3 Factors Influencing the Manifestation of Audiovisual Integration Effects

One strength of our study was that we simultaneously found both audiovisual sub-additivity and enhancement, highlighting the complex dynamics of multisensory integration. We suggest that whether multisensory integration is reflected in enhancement or sub-additivity depends largely on the degree to which unisensory processes recruit a region. In regions in which unimodal stimuli evoke robust activation, only sub-additivity is likely to be detected. Conversely, enhancement effects are most likely to appear in regions only weakly activated by unimodal stimuli. Both enhancement and sub-additivity are evidence of multisensory interactions, and both represent a modification of neural representational codes. At the spatial scale of individual voxels, these modifications manifest differently, however at the neuronal level, these effects may involve identical mechanisms (i.e., inhibiting some populations and exciting others, but in different proportions or over different time windows).

The enhancement effects observed in pSTS were consistent with previous findings of audiovisual additivity in this region during language processing (Stevenson & James, 2009; van Atteveldt et al., 2004). One implication of the inverse effectiveness principle is that regions most active for unimodal stimuli (e.g., showing the greatest difference when contrasting auditory- versus visual-only) would be most likely to show sub-additivity. Previous work (Binder et al., 2006; Graves, Grabowski, Mehta, & Gupta, 2008; McNorgan, Awati, et al., 2013) lead us to predict that sub-additivity would be seen in left STG, and left FG, as primary phonological and orthographic processing regions, respectively. The failure to find audiovisual sub-additivity in STG was inconsistent our prediction and with previous findings showing visual suppression in primary auditory processing areas (Hertrich,

Dietrich, & Ackermann, 2010; Lehmann et al., 2006). As a null effect, however, this provides little insight on its own, and as we suggested earlier, may be attributable to the phonological demands of the experimental task. The predicted sub-additivity for FG was in line with previous findings of phonological suppression of activity in FG (Yoncheva et al., 2010), though as we indicated earlier, the AV condition in our experiment was associated with a phonological modulation of activity towards baseline, and not a suppression of activity below baseline.

We interpreted the observed sub-additive effects in the context of two behavioral correlates (reading skill and reaction time). Behaviorally, the AA condition was associated with longer reaction times compared to the other two presentation modalities. Reaction time is frequently used as an index of task difficulty (Mckiernan et al., 2003), and thus this difference could be taken as an indicator of increased processing difficulty for the AA condition. As we noted earlier, however, this likely also reflects the fact that rhyming judgments in the AA condition cannot be made before the item has been enunciated to completion. Because we used response time as a dependent measure, we did not factor out the variance associated with this variable in our analyses, using instead accuracy as a measure of difficulty. This provides mechanistic and cognitive constructs (e.g., learning, priming) in which our neuroimaging results can be grounded, and provides insight into how audiovisual integration depends on task and participant characteristics. Thus, the potential difficulty confound introduced by the significant main effect of response time across task is a potential limitation of our study, though we hope to have largely mitigated the problem by accounting for response accuracy, which has also been used as a behavioral index of task difficulty by investigators explicitly looking at the relationship between the BOLD signal and task difficulty (e.g., Gould et al., 2003; Mckiernan et al., 2003).

Writing is a relatively recent cultural invention and the audiovisual integration processes underlying reading therefore presumably depend heavily on experience, yet few neuroimaging studies have examined skill-related audiovisual integration during reading, and most of these have focused on letter-sound integration in adults. Though previous research has demonstrated skill-related cross-modal enhancement and sub-additivity differentiates typical and dyslexic adults (Blau, van Atteveldt, Ekkebus, Goebel, & Blomert, 2009), these effects were observed only in posterior superior temporal cortex. Audiovisual integration of letters and sounds in the visual processing stream in adults has been shown to depend on stimulus characteristics (Blau et al., 2008), again leaving open the question of how learning to read shapes the development of the orthographic processing system. In children, elision, a commonly-used measure of phonemic awareness, appears to depend on audiovisual integration in FG for typically developing readers but not those with reading difficulty (McNorgan, Randazzo-Wagner, & Booth, 2013), suggesting that fluent reading depends on audiovisual integration in the orthographic system. Our results are consistent with, and extend this literature by showing that reading ability is related to the degree to which audiovisual integration processes modulate activity in the visual-object processing stream, thereby facilitating recognition of printed words.

5. Conclusion

Our study provides the clearest evidence to date of phonological modulation of the orthographic processing stream, including the putative visual word form area, and provides critical support for the interactive account (Price & Devlin, 2011). That this relationship exists along the length of the orthographic processing gradient suggests that a developmental change in the nature of audiovisual integration along the visual-object processing stream helps optimize this region for recognizing printed language at all levels of orthographic grain size. Overall, these results further reinforce the dependency of heteromodal modulation and multisensory integration effects on the inverse effectiveness of the unimodal stimuli.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants from the National Institute of Child Health and Human Development (HD042049) to James R. Booth.

References

- Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, Loftis B, Treiman R. The English Lexicon Project. *Behavior Research Methods*. 2007; 39(3):445–459. [PubMed: 17958156]
- Balota DA, Yap MJ, Cortese MJ, Watson JM. Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*. 2008; 59(4):495–523. doi: <http://dx.doi.org/10.1016/j.jml.2007.10.004>.
- Beauchamp MS. Statistical Criteria in fMRI Studies of Multisensory Integration. *Neuroinformatics*. 2005; 3(2):093–114.10.1385/ni:3:2:093
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron*. 2004; 41(5):809–823.10.1016/s0896-6273(04)00070-4 [PubMed: 15003179]
- Bernstein LE, Auer ET Jr, Wagner M, Ponton CW. Spatiotemporal dynamics of audiovisual speech processing. *NeuroImage*. 2008; 39(1):423–435. doi: <http://dx.doi.org/10.1016/j.neuroimage.2007.08.035>. [PubMed: 17920933]
- Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*. 2006; 33(2):739–748. [PubMed: 16956773]
- Bishop CW, Miller LM. A Multisensory Cortical Network for Understanding Speech in Noise. *Journal of Cognitive Neuroscience*. 2008; 21(9):1790–1804.10.1162/jocn.2009.21118 [PubMed: 18823249]
- Blau V, van Atteveldt NM, Ekkebus M, Goebel R, Blomert L. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. *Curr Biol*. 2009; 19(6): 503–508.10.1016/j.cub.2009.01.065 [PubMed: 19285401]
- Blau V, van Atteveldt NM, Formisano E, Goebel R, Blomert L. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur J Neurosci*. 2008; 28(3):500–509.10.1111/j.1460-9568.2008.06350.x [PubMed: 18702722]
- Brem S, Bach S, Kucian K, Guttorm TK, Martin E, Lyytinen H, Richardson U. Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*. 2010; 107(17):7939–7944.
- Buckner RL, Goodman J, Burock M, Rotte M, Koutstaal W, Schacter D, Dale AM. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*. 1998; 20(2):285–296. [PubMed: 9491989]

- Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*. 2000; 10(11):649–658. [PubMed: 10837246]
- Cohen L, Jobert A, Le Bihan D, Dehaene S. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *Neuroimage*. 2004; 23(4):1256–1270.10.1016/j.neuroimage.2004.07.052 [PubMed: 15589091]
- Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S. Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*. 2002; 125(5):1054–1069. [PubMed: 11960895]
- Dehaene S, Cohen L, Sigman M, Vinckier F. The neural code for written words: a proposal. *Trends in Cognitive Sciences*. 2005; 9(7):335–341.10.1016/j.tics.2005.05.004 [PubMed: 15951224]
- Goebel R, van Atteveldt N. Multisensory functional magnetic resonance imaging: a future perspective. *Experimental brain research*. 2009; 198(2–3):153–164. [PubMed: 19533111]
- Goswami U. Causal connections in beginning reading: the importance of rhyme. *Journal of Research in Reading*. 1999; 22(3):217–240.10.1111/1467-9817.00087
- Gould RL, Brown RG, Owen AM, ffytche DH, Howard RJ. fMRI BOLD response to increasing task difficulty during successful paired associates learning. *NeuroImage*. 2003; 20(2):1006–1019. doi: [http://dx.doi.org/10.1016/S1053-8119\(03\)00365-3](http://dx.doi.org/10.1016/S1053-8119(03)00365-3). [PubMed: 14568470]
- Graves WW, Grabowski TJ, Mehta S, Gupta P. The Left Posterior Superior Temporal Gyrus Participates Specifically in Accessing Lexical Phonology. *Journal of Cognitive Neuroscience*. 2008; 20(9):1698–1710.10.1162/jocn.2008.20113 [PubMed: 18345989]
- Grill-Spector K, Henson R, Martin A. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*. 2006; 10(1):14–23. [PubMed: 16321563]
- Hertrich I, Dietrich S, Ackermann H. Cross-modal Interactions during Perception of Audiovisual Speech and Nonspeech Signals: An fMRI Study. *Journal of Cognitive Neuroscience*. 2010; 23(1):221–237.10.1162/jocn.2010.21421 [PubMed: 20044895]
- Klemen J, Chambers CD. Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience and Biobehavioral Reviews*. 2012; 36(1):111–133. doi: <http://dx.doi.org/10.1016/j.neubiorev.2011.04.015>. [PubMed: 21569794]
- Knecht S, Dräger B, Deppe M, Bobe L, Lohmann H, Flöel A, Henningsen H. Handedness and hemispheric language dominance in healthy humans. *Brain*. 2000; 123(12):2512–2518.10.1093/brain/123.12.2512 [PubMed: 11099452]
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*. 2005; 166(3–4):289–297. [PubMed: 15988597]
- Lehmann C, Herdener M, Esposito F, Hubl D, di Salle F, Scheffler K, Seifritz E. Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *NeuroImage*. 2006; 31(1):294–300. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.12.038>. [PubMed: 16473022]
- Los SA, Van der Burg E. Sound speeds vision through preparation, not integration. *Journal of Experimental Psychology: Human Perception and Performance*. 2013; 39(6):1612. [PubMed: 23477697]
- Mckiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of cognitive neuroscience*. 2003; 15(3):394–408. [PubMed: 12729491]
- McNorgan C, Alvarez A, Bhullar A, Gayda J, Booth JR. Prediction of Reading Skill Several Years Later Depends on Age and Brain Region: Implications for Developmental Models of Reading. *The Journal of Neuroscience*. 2011; 31(26):9641–9648.10.1523/jneurosci.0334-11.2011 [PubMed: 21715629]
- McNorgan C, Awati N, Desroches AS, Booth JR. Multimodal Lexical Processing in Auditory Cortex Is Literacy Skill Dependent. *Cerebral Cortex*. 2013; 23(10):1093–1100.10.1093/cercor/bht100
- McNorgan C, Randazzo-Wagner M, Booth JR. Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Frontiers in Human Neuroscience*. 2013; 7:10.3389/fnhum.2013.00388 [PubMed: 23372547]

- Noppeney, U. Characterization of multisensory integration with fMRI: experimental design, statistical analysis, and interpretation. Boca Raton, FL: CRC Press; 2012.
- Price CJ, Devlin JT. The myth of the visual word form area. *Neuroimage*. 2003; 19(3):473–481. [PubMed: 12880781]
- Price CJ, Devlin JT. The interactive account of ventral occipitotemporal contributions to reading. [Research Support, Non-U.S. Gov't]. *Trends in Cognitive Sciences*. 2011; 15(6):246–253.10.1016/j.tics.2011.04.001 [PubMed: 21549634]
- Pugh KR, Frost SJ, DSandak R, Landi N, Moore D, Della Porta G, Pugh K. Mapping the word reading circuitry in skilled and disabled readers. *The neural basis of reading*. 2010:281–305.
- Schlaggar BL, McCandliss BD. Development of neural systems for reading. *Annual Review of Neuroscience*. 2007; 30:475–503.
- Scott SK, Rosen S, Wickham L, Wise RJS. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *The Journal of the Acoustical Society of America*. 2004; 115(2):813–821. doi:<http://dx.doi.org/10.1121/1.1639336>. [PubMed: 15000192]
- Share DL. Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*. 1995; 55(2):151–218. [PubMed: 7789090]
- Stein BE, Stanford TR. Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*. 2008; 9(4):255–266.10.1038/Nrn2331
- Stevenson RA, James TW. Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*. 2009; 44(3):1210–1223.10.1016/j.neuroimage.2008.09.034 [PubMed: 18973818]
- Stevenson RA, Wallace MT. Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*. 2013:1–13.
- Sugihara T, Diltz MD, Averbek BB, Romanski LM. Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. *The Journal of Neuroscience*. 2006; 26(43):11138–11147.10.1523/jneurosci.3550-06.2006 [PubMed: 17065454]
- Szaflarski JP, Binder JR, Possing ET, McKiernan KA, Ward B, Hammeke TA. Language lateralization in left-handed and ambidextrous people fMRI data. *Neurology*. 2002; 59(2):238–244. [PubMed: 12136064]
- Thompson-Schill SL, D'Esposito M, Kan IP. Effects of Repetition and Competition on Activity in Left Prefrontal Cortex during Word Generation. *Neuron*. 1999; 23(3):513–522. doi: [http://dx.doi.org/10.1016/S0896-6273\(00\)80804-1](http://dx.doi.org/10.1016/S0896-6273(00)80804-1). [PubMed: 10433263]
- Tsapkini K, Rapp B. The orthography-specific functions of the left fusiform gyrus: Evidence of modality and category specificity. *Cortex*. 2010; 46(2):185–205. doi: <http://dx.doi.org/10.1016/j.cortex.2009.02.025>. [PubMed: 19428003]
- Turkeltaub PE, Coslett HB. Localization of sublexical speech perception components. *Brain and Language*. 2010; 114(1):1–15. doi: <http://dx.doi.org/10.1016/j.bandl.2010.03.008>. [PubMed: 20413149]
- Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation. *Neuroimage*. 2002; 16(3 Part A):765–780.10.1006/nimg.2002.1131 [PubMed: 12169260]
- Tyll S, Bonath B, Schoenfeld MA, Heinze HJ, Ohl FW, Noesselt T. Neural basis of multisensory looming signals. *NeuroImage*. 2013; 65(0):13–22. doi: <http://dx.doi.org/10.1016/j.neuroimage.2012.09.056>. [PubMed: 23032489]
- van Atteveldt NM, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex*. 2007; 17(4):962–974.10.1093/cercor/bhl007 [PubMed: 16751298]
- van Atteveldt NM, Formisano E, Goebel R, Blomert L. Integration of letters and speech sounds in the human brain. *Neuron*. 2004; 43(2):271–282.10.1016/j.neuron.2004.06.025 [PubMed: 15260962]
- van Atteveldt NM, Peterson BS, Schroeder CE. Contextual control of audiovisual integration in low-level sensory cortices. *Human Brain Mapping*. 2014; 35(5):2394–2411.10.1002/hbm.22336 [PubMed: 23982946]

- van Atteveldt NM, Roebroek A, Goebel R. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. *Hear Res.* 2009; 258(1–2):152–164.10.1016/j.heares.2009.05.007 [PubMed: 19500658]
- Vigneau M, Beaucoisin V, Hervé PY, Jobard G, Petit L, Crivello F, Tzourio-Mazoyer N. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: Insights from a meta-analysis. *NeuroImage.* 2011; 54(1):577–593. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.07.036>. [PubMed: 20656040]
- Wiggs CL, Martin A. Properties and mechanisms of perceptual priming. *Current opinion in neurobiology.* 1998; 8(2):227–233. [PubMed: 9635206]
- Wong PC, Uppunda AK, Parrish TB, Dhar S. Cortical mechanisms of speech perception in noise. *Journal of Speech, Language, and Hearing Research.* 2008; 51(4):1026–1041.
- Yoncheva YN, Zevin JD, Maurer U, McCandliss BD. Auditory Selective Attention to Speech Modulates Activity in the Visual Word Form Area. *Cerebral Cortex.* 2010; 20(3):622–632.10.1093/cercor/bhp129 [PubMed: 19571269]
- Zeno, S.; Ivens, S.; Millard, R.; Duvvuri, R. The educator's word frequency guide/Touchstone Applied Science Associates. Brewster, NY: Touchstone Applied Science Associates; 1995.

Highlights

1. Explored skill-dependent changes in audiovisual (AV) integration in fusiform gyrus
2. Provide first evidence of theoretically-predicted AV suppression in this region
3. Greater reading skill predicted greater AV suppression
4. Greater AV suppression indicated greater crossmodal orthographic priming
5. Learning to read promotes phonological facilitation on orthographic processing

Figure 1. Schematic diagram of the task design for the audiovisual (a), and unimodal auditory-only (b) and visual-only (c) task conditions. Stimulus presentation timing was identical across task conditions, and trial onsets were jittered to fall between 2200 and 2800 ms apart.

Figure 2.

Whole-brain renderings showing overall [lexical > fixation] contrast activation (warm colors) and [fixation > lexical] contrast deactivation (cool colors) for the audiovisual (a) and unimodal auditory-only (b) and visual-only (c) task conditions. Deactivations were inclusively masked to show only deactivation in regions also showing activation in any of the three modality conditions.

Figure 3.

Whole-brain rendering of super-additive audiovisual effects (a) and additive audiovisual effects (b) show regions demonstrating an enhanced hemodynamic response to the cross-modal condition compared to both unimodal conditions. Hemispheric differences in additivity effects were explored within left- and right-hemisphere regions of interest (c) comprised of the intersection of voxels showing additive effects or super-additive effects with the contralateral mirror of these voxels in the left (green), and right (blue) hemispheres.

Figure 4.

Laterality differences in additive and super-additive audiovisual enhancement were attributable to the inverse effectiveness principle. Left-hemisphere audiovisual enhancement was additively greater than the unimodal conditions. Right-hemisphere superadditive audiovisual enhancement was driven by reduced activation for the unimodal conditions relative to baseline.

Figure 5. Whole-brain rendering of regions demonstrating reading-skill dependent sub-additivity of the hemodynamic response to the cross-modal condition compared to both unimodal conditions.

Figure 6.

Increased reading skill (Word ID) was associated with decreased audiovisual activation for the cross-modal task in both middle occipital gyrus (A) and fusiform gyrus (B) regions of interest, relative to both unimodal conditions. A significant positive correlation was found between behavioral priming effects and audiovisual suppression in middle occipital gyrus (C) and fusiform gyrus (D). Orthographic priming effects for the cross-modal condition are plotted along the y-axis. Positive values indicate greater priming effects, i.e. faster responses for matching versus mismatching orthography. Audiovisual suppression (i.e. sub-additivity) is plotted along the x-axis. Positive values indicate greater suppression of the cross-modal relative to the unimodal condition

Table 1

Mean stimulus word length (letters), word frequency and offline naming accuracy by congruency condition (standard deviations in parentheses).

Condition	Word Length	Frequency ¹	Naming Accuracy ²
O+P+	4.5 (0.6)	5.8 (6.5)	1.00 (.01)
O+P-	4.7 (0.8)	11.9 (14.5)	.99 (.01)
O-P+	4.5 (0.6)	7.7 (11.0)	.99 (.01)
O-P-	4.6 (0.6)	10.8 (12.3)	.99 (.01)

¹ Occurrences per million words; *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995);

² *The English Lexicon Project* (Balota et al., 2007)

Table 2

Mean accuracy and decision latencies for correct trials (in milliseconds) for orthographically similar (O+P+) and dissimilar (O-P+) rhyming items by task modality. Standard errors appear in parentheses.

		Condition	
		O+P+	O-P+
Accuracy	AV	0.80 (0.02)	0.82 (0.02)
	AA	0.78 (0.03)	0.86 (0.02)
	VV	0.94 (0.01)	0.91 (0.02)
Decision Latency	AV	1086 (50)	1102 (57)
	AA	1425 (61)	1370 (54)
	VV	1194 (50)	1259 (55)

Note: Decision latencies reported in milliseconds

Table 3

Coordinates for clusters reaching $q < 0.01$ (FDR cluster-level corrected) significance in the lexical minus fixation (Lex>Fix) and masked fixation minus lexical (Fix>Lex) contrasts for words for the AV, AA and VV modality conditions

Modality Contrast	Region	BA	k	x	y	z	Z	FDR
AV Lex>Fix	L STG/MTG	21/22/48	6695	-54	-18	6	6.72	<.001
	R STG/MTG	22/21/48	2278	64	-20	0	6.48	<.001
	L/R SMA/Anterior Cingulate	6/32	1521	-4	18	44	6.27	<.001
	L Fusiform Gyrus	19	1466	-42	-66	-12	5.62	<.001
	L Calcarine Sulcus	18	5003	0	-84	16	5.50	<.001
	R Inferior Occipital Gyrus	19	394	48	-78	-2	5.13	<.001
	R Insula	47	417	32	28	0	4.90	<.001
	L Thalamus		150	-8	-14	4	4.76	<.001
	L Superior Parietal Lobule	7	321	-26	-66	-50	4.51	<.001
No significant clusters found								
Fix>Lex								
AA Lex>Fix	L Heschl's Gyrus	21/22/48	4109	-38	-30	12	7.15	<.001
	R STG	21/22	2471	58	-24	0	6.95	<.001
	L SMA/Anterior Cingulate	6/32	346	-8	14	50	5.62	<.001
	R Cerebellum		105	24	-64	-24	4.94	<.001
	R Cuneus	18	533	14	-86	22	4.83	<.001
	L IFG (Pars Operculus)	48/45	100	-46	16	16	4.07	<.001
	L Calcarine Sulcus	17	112	-10	-78	10	4.04	<.001
	L Inferior Occipital Gyrus	19	172	-36	-86	-8	4.57	<.001
Fix>Lex								
VV Lex>Fix	L/R SMA/Anterior Cingulate/SMA	6/32/24	861	-2	18	48	6.69	<.001
	L Hippocampus/Thalamus		332	-22	-28	-4	6.41	<.001
	L Lingual Gyrus	18	3213	-22	-92	-14	6.22	<.001
	R Fusiform Gyrus/Cerebellum	18/19	5834	28	-70	-24	6.09	<.001
	L Precentral/Postcentral Gyrus/ IFG (Pars Opercularis)	6/4/44	4361	-50	0	46	5.99	<.001
	L MTG	21/22	265	-60	-36	2	5.36	<.001
	R Precentral/Postcentral Gyrus	6/4	115	56	4	42	5.33	<.001
	R Hippocampus		84	24	-24	-6	5.19	<.001
	R Lingual Gyrus	27	52	16	-44	4	5.17	.003

Modality Contrast	Region	BA	k	x	y	z	Z	FDR
Fix>Lex	L Putamen		56	-18	-4	-10	4.77	.002
	R Hippocampus		50	34	-30	-6	4.50	.004
	No significant clusters found							

L, left; R, right; IFG, Inferior Frontal Gyrus.; STG, Superior Temporal Gyrus; MTG, Middle Temporal Gyrus; SMA, Supplementary Motor Area; BA, Brodmann Area; k, cluster spatial extent in 8mm³ voxels; Z, peak Z-score; FDR, false discovery rate cluster-extent corrected significance. Coordinates reflect standard MNI space

Table 4
Audiovisual enhancement of the hemodynamic response, controlling for task performance and reading skill.

Audiovisual Enhancement	Region	BA	k	x	y	z	Z	FDR
Additivity	L IFG	44/9	612	-54	18	24	4.64	<.001
	L Orbitofrontal Cortex	47	67	-28	28	2	4.02	.025
	L MTG	21	65	-60	-44	-2	3.49	.025
	L IPL	40	150	-32	-58	40	3.44	<.001
Super-additivity	R IPL	40	211	32	-48	42	4.46	<.001
	R IFG	44/9	377	40	6	32	4.23	<.001
	R IPL/STG/MTG	40/22/21	429	58	-38	36	4.14	<.001
	R Caudate		87	10	6	14	3.92	.011

L, left; R, right; IFG, Inferior Frontal Gyrus.; IPL, Inferior Parietal Lobule; STG, Superior Temporal Gyrus; MTG, Middle Temporal Gyrus; BA, Brodmann Area; k, cluster spatial extent in 8mm³ voxels; Z, peak Z-score; FDR, false discovery rate cluster-extent corrected significance. Coordinates reflect standard MNI space.

Table 5
Reading skill-dependent audiovisual sub-additivity of the hemodynamic response, controlling for task performance.

Region	BA	k	x	y	z	Z	FDR
L Fusiform Gyrus	19/37	265	-32	-68	-14	4.50	<.001
L Middle Occipital Gyrus	19	104	-50	-72	4	3.99	.010
R Middle Frontal Gyrus	46	177	42	52	6	3.97	<.001

L, left; R, right; BA, Brodmann Area; k, cluster spatial extent in 8mm³ voxels; Z, peak Z-score; FDR, false discovery rate cluster-extent corrected significance. Coordinates reflect standard MNI space.