Cerebral Cortex September 2014;24:2464–2475
doi:10.1093/cercor/bht100
Advance Access publication April 15, 2013

Multimodal Lexical Processing in Auditory Cortex Is Literacy Skill Dependent

Chris McNorgan, Neha Awati, Amy S. Desroches and James R. Booth

Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208, USA

Address correspondence to Dr Chris McNorgan. Email: chris.mcnorgan@alumni.uwo.ca

Literacy is a uniquely human cross-modal cognitive process wherein visual orthographic representations become associated with auditory phonological representations through experience. Developmental studies provide insight into how experience-dependent changes in brain organization influence phonological processing as a function of literacy. Previous investigations show a synchrony-dependent influence of letter presentation on individual phoneme processing in superior temporal sulcus; others demonstrate recruitment of primary and associative auditory cortex during cross-modal processing. We sought to determine whether brain regions supporting phonological processing of larger lexical units (monosyllabic words) over larger time windows is sensitive to cross-modal information, and whether such effects are literacy dependent. Twenty-two children (age 8–14 years) made rhyming judgments for sequentially presented word and pseudoword pairs presented either unimodally (auditory- or visual-only) or cross-modally (audiovisual). Regression analyses examined the relationship between literacy and congruency effects (overlapping orthography and phonology vs. overlapping phonology-only). We extend previous findings by showing that higher literacy is correlated with greater congruency effects in auditory cortex (i.e., planum temporale) only for cross-modal processing. These skill effects were specific to known words and occurred over a large time window, suggesting that multimodal integration in posterior auditory cortex is critical for fluent reading.

Keywords: audiovisual integration, cross-modal, development, fMRI, reading

Introduction

Multimodal processing is pervasive in cognition, and therefore understanding how the brain integrates multimodal information is critical to understanding cognitive processing in general. Reading—a process in which a learned set of arbitrary visual symbols are mapped on to auditory phonological representations—is a particularly interesting multimodal process because it requires explicit effort to learn these mappings and because it is performed with a wide range of ability. The systems that integrate visual and auditory representations provide the foundation for the uniquely human capability for language. Understanding language development therefore requires an appreciation of its reliance on the development of the system underlying audiovisual integration. Contemporary theories of language development hold that early exposure to spoken language teaches infants both the allowable set of phonemes in their native language (Kuhl 2004), and the corresponding articulatory patterns required for production (Liberman and Mattingly 1985). A clear example of multimodal integration in language processing can be found in the domain of speech perception. The McGurk effect (McGurk and Macdonald 1976) describes a phenomenon in which phonological perception of an utterance is influenced by a mismatching visual perception of the speaker's mouth during articulation. Thus, knowledge about word articulation additionally helps disambiguate or alter the perception of speech by providing visual cues that augment noisy speech signals.

Cross-Modal Congruency

Stimulus congruency is an important tool in the investigation of cross-modal interaction. We learn by experience that some objects, events, or stimulus characteristics typically co-occur in a particular context. Congruent pairings meet these expectations, whereas incongruent pairings violate them. If responses to congruent and incongruent pairings differ, this suggests that the processing of one stimulus influences the processing of the other. The McGurk effect, for example, arises from a mismatch between the visual shape of the speaker's mouth and the corresponding phoneme. Investigations of cross-modal processing assess congruency effects between modalities to determine the extent to which one modality influences another (e.g., Beauchamp et al. 2004).

The literature pointing to the posterior superior temporal sulculs (pSTS) as a brain region critical for audiovisual integration has done so largely on the weight of evidence provided by sensitivity of this region to congruency between auditory and visual input in multiple contexts, including spoken language processing and object recognition (Koelewijn et al. 2010). This research suggests that pSTS plays a role in audiovisual integration more generally, although within the pSTS, language-specific cross-modal integration sites may coexist with other subpopulations involved in cross-modal audiovisual integration in other domains (Hein and Knight 2008; Stevenson and James 2009). Spoken language processing, however, is an early developing skill. Moreover, although it may be influenced by visual information, spoken language processing does not depend on it. In contrast, the mapping of orthographic to phonologic representations during reading is intrinsically cross-modal.

Previous Developmental Neuroimaging Studies of Cross-Modal Linguistic Processing

Adult studies comprise the vast majority of the neuroscience literature on audiovisual integration. Models of brain-behavior correlations drawn from studies of skilled adult performance, however, may rely on incomplete or inaccurate assumptions about how these skills emerge. Language acquisition is an interesting model process for studying the relationship between brain organization and cognitive skill because, although an infant's sensitivity to her native language (Dehaene-Lambertz et al. 2002) suggests the brain is predisposed toward learning language, reading fluency emerges only after extensive and explicit instruction for most, and does so inadequately for a

large segment of the population experiencing reading difficulty (Shaywitz et al. 1990). A developmental cognitive neuroscience approach thus provides important insight into the functional brain organization underlying typical and atypical reading (Schlaggar and McCandliss 2007).

Letters and phonemes constitute the basic elements over which written and spoken language are associated in alphabetic languages, and learning these associations is a necessary precursor to literacy (Frith 1985). Accordingly, a number of researchers have conducted investigations of cross-modal integration of orthographic and phonologic representations at the level of individual letters and phonemes. Audiovisual integration has been shown to be synchrony dependent (Miller and D'Esposito 2005; van Atteveldt, Formisano, Blomert et al. 2007). This sensitivity to the temporal characteristics of the stimuli makes event-related potentials (ERPs), which have millisecond-level temporal resolution, an excellent tool for investigating the development of the system underlying audiovisual integration. In ERP studies of multimodal letter-speech processing, early responses differed for letter-phoneme pairs in the audiovisual condition and those for the auditory-only condition, but these differences were apparent only for advanced (fifth-grade) readers but not less-skilled (second-grade) readers (Froyen, Bonte et al. 2008). Interestingly, this modality difference was observed in the older children only when letters were presented 200 ms before the corresponding auditory stimulus; a similar study using adults found these effects only when the paired audiovisual stimuli were presented simultaneously (Froyen, van Atteveldt et al. 2008). Together, these results indicate that rapid automatic audiovisual integration of letters and their corresponding phonemic representations is a developing skill.

A complete picture of the development of language-related audiovisual integration additionally requires understanding how the neural substrates underlying audiovisual processing behave as a function of language experience. Much of what is known about the localization of these processes comes from functional magnetic resonance imaging (fMRI) studies, which point to pSTS as a region critical for audiovisual integration (Nath and Beauchamp 2012). Within children, activity within left pSTS predicts susceptibility to the McGurk illusion (Nath et al. 2011), indicating that this region is recruited during audiovisual integration in speech perception, although developmental changes in other nearby left-hemisphere perisylvian structures have also been implicated in the maturation of the system underlying audiovisual integration during both lexical and speech processing (Booth et al. 2004; Ross et al. 2011). Taken as a whole, the current body of neuroscience literature thus suggests that automaticity of audiovisual integration during language processing is accompanied by developmental changes within left posterior temporal and temporoparietal cortex.

Motivation and Predictions

van Atteveldt and coworkers (van Atteveldt, Formisano, Blomert et al. 2007; van Atteveldt, Formisano, Goebel et al. 2007) used the phoneme-letter congruency effect as an indicator of multisensory integration because the congruency of a pair of stimuli is a property not of either of the items individually, but instead of the product of their integration. In the present experiment, we manipulate orthographic rime-level congruency in children spanning a range of ages performing a rhyming decision task to explore the development of cross-modal processing in reading. Our position is that developmental changes in cross-modal processing arise from experience with written and spoken language. Accordingly, we are primarily interested in how cross-modal processing, as measured by sensitivity to interstimulus congruency, changes as a function of literacy.

The mapping of individual letters to their corresponding speech sounds mirrors the explicit instruction in the alphabetic principle that many children receive. Even before this instruction, however, a child will have acquired a substantial spoken language vocabulary comprised of word-level tokens. Thus, even beginning readers operate on lexical units at various grain sizes (or granularity) (Ziegler and Goswami 2005). Grain size refers to the size of the unit over which a lexical system (e.g., phonological or orthographic processes) operates. Lexical objects can be considered with respect to a continuum of granularity. Whole words exist at the largest phonological and orthographic grain sizes. Syllables (phonological) and *n*-grams (orthographic) are units at an intermediate granularity; phonemes (phonological) and graphemes/letters (orthographic) are units at an even smaller granularity, and sensitivity to these different levels of granularity changes developmentally (Ziegler and Goswami 2005). It remains unclear whether the integration effects observed using single letters and phonemes apply at the whole-word level. If they do, one implication is that phonological representations generated during reading are continuously updated by cross-modal information, rather than being an encapsulated product of item-by-item integration at a more atomic grapheme/phoneme level. A corollary of this is that the requirement of temporal synchrony required for cross-modal integration of individual letters and phonemes (van Atteveldt, Formisano, Blomert et al. 2007) is relaxed for larger grain sizes.

Following the logic of previous neuroimaging investigations, we assume that cross-modal congruency effects are an index of multimodal processing of written lexical stimuli. As discussed earlier, pSTS has been implicated as an audiovisual integration area (for a review, see Calvert 2001). This literature, however, is largely based on studies of online integration of perisynchronously presented stimuli at the smallest grain size (i.e., phonemes/graphemes). A critical aspect of fluent reading is that phonological representations are internally generated and are driven by orthographic representations. That is, when one reads a word, the evoked phonological representation comes not from an external auditory source. Rather, the representation emerges over time, driven by the association one has learned between the visually perceived word form and the corresponding phonological word form. It is reasonable to assume, therefore, that cross-modal influences at larger grain sizes would be apparent in regions that encode and/or maintain phonological representations. As described earlier, there is conditional evidence for cross-modal integration in auditory cortex (Heschl's gyrus and planum temporale; HG and PT) (Calvert 1997; Hein et al. 2007; Hickok et al. 2009). Importantly, the PT seems to participate in the phonological store (Buchsbaum and D'Esposito 2008), which suggests that this region may play an important role in integrating lexical stimuli at larger grain sizes because these lexical representations unfold over time as they are

assembled from smaller units. If the network that carries out this integration develops through experience with written language, the sensitivity of these regions to cross-modal congruency should be modulated by literacy: Those with higher skill should be more sensitive to cross-modal information. Accordingly, we predict a significant positive correlation between literacy measures and the fMRI congruency effect in auditory cortex for cross-modally presented stimuli. The body of literature reviewed here tends to show evidence for audiovisual integration in more posterior temporal regions. Thus, congruency effects should strongest in posterior PT or pSTS. We therefore expect correlations between literacy and congruency effects should be strongest in these regions.

Materials and Methods

Participants

A group of 22 typically achieving children (12 males; mean age = 10.92 years, standard deviation (SD) = 1.5 years, range = 8.58–13.58 years) participated in the present study. All participants were native English speakers, right handed, had normal or corrected-to-normal vision, normal hearing, and had no history of psychiatric illness, neurological disease, learning disability, or attention deficit hyperactivity disorder. Participants were recruited from the Chicago metropolitan area. Informed consent was obtained from participants and their parents, and all procedures were approved by the Institutional Review Board at Northwestern University.

Prior to admission to the study, we evaluated children's verbal and nonverbal intelligence quotient (IQ) using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler 1999). For inclusion in the present study, each child's full scale (i.e., verbal and nonverbal IQ score) was at or above a standard score of 100, with verbal IQ being >95 (Verbal IQ: M = 123, SD = 14; nonverbal IQ: M = 115, SD = 12). Reading and spelling subtests of the Woodcock-Johnson-III Tests of Achievement (WJ-III; Woodcock et al. 2001) were used to assess 2 aspects of literacy. The reading subtest measures the fluency with which an individual is able to decode a textual statement (e.g., "milk is pink") for meaning. The spelling subtest measures the accuracy with which an individual is able to spell a spoken word (i.e., map from phonology to orthography).

Experimental Procedure

Rhyme Judgment Task

On each 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 and visual modalities, respectively. Previous investigations of cross-modal lexical processing research (e.g., van Atteveldt et al. 2004; van Atteveldt, Formisano, Goebel et al. 2007; Froyen, Bonte et al. 2008) similarly employed auditory-then-visual presentations, motivating the task design for that modality condition. 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 their right index finger for a yes (rhyme) response and their right middle finger for a no (nonrhyme) response. The rhyming task requires phonological retrieval and maintenance. To act as a comparison target for the second stimulus, the initially presented item must be maintained in memory and, in the case of the AV and VV conditions, the phonological representations of one or both stimuli must be accessed from the orthographic representation. This task thus seemed ideally suited for investigating cross-modal processing within phonological processing areas.

Although we were primarily interested in performance for word stimuli, participants were additionally administered a pseudoword condition. The inclusion of a pseudoword condition provided an alternative measure of experience-dependent effects—that is, whether familiarity with a word form influences multisensory integration. We thus included these data in our analyses to increase the confidence with which we could argue that our results were driven by reading experience. Two word and pseudoword runs for each modality condition were presented in separate runs each lasting ~7 min. Thus, the total duration of the word and pseudoword runs for each scanning session was ~28 min. Each stimulus item was presented for 800 ms, separated by a 200-ms interstimulus interval. 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 participant 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, although this accounted for only 13% of all errors and occurred on 4.1% of the trials. Similarly, responses made before the second stimulus had been fully presented occurred on <0.2 percent of the trials. Thus, children had little difficulty in responding within the time window provided. A jittered response interval duration of between 2200 and 2800 ms was used to allow for deconvolution of the signal associated with each condition. The sequence and timings of lexical trial events are illustrated for each modality in Figure 1.

Stimulus pairs varied in terms of their orthographic and phonological similarity, and were presented in 1 of 4 similarity conditions (24 pairs per condition, per lexicality). There were 2 phonologically similar (i.e., rhyming) conditions, one with orthographically similar pairs (O+P+) and another with orthographically dissimilar pairs (O−P+). There were also 2 phonologically dissimilar (i.e., nonrhyming) conditions, 1 with orthographically similar pairs (O+P−) and 1 with orthographically dissimilar pairs (O−P−). Example rhyming word and pseudoword pairs are presented in Table 1. All words were monosyllabic, having neither homophones nor homographs. Pseudowords were adapted from real words by replacing the initial consonant(s) of a real word to make a novel item. We verified that no pseudoword was a real word or pseudo-homophone using an online dictionary (http://dictionary.reference.com). All words were matched across conditions for written word frequency in children (Zeno 1995) and the sum of their written bigram frequency (English Lexicon Project, http://elexicon.wustl.edu).

For the fMRI analyses, a number of considerations led to the restriction of the analyses to the 2 rhyming conditions. First, we were primarily interested in the effect of orthographic congruency, for which there were 2 rhyming conditions at each level of orthographic congruency (e.g., rhyming congruent, O+P+; and nonrhyming congruent, O+P−). It is likely that the "no" responses in the nonrhyming conditions engage the phonological network differently than the "yes" responses. Collapsing across rhyming and nonrhyming conditions would have necessitated analyses of the interaction between congruency and rhyme (for which we had no predictions) in order to isolate the effect of congruency. A second consideration concerned the pronunciation of rhyming versus nonrhyming pseudowords, for which pronunciation is ambiguous. Statistical properties of the English language determined whether pseudoword pairs rhymed, probabilistically. Nonetheless, rhyming pronunciations of incongruent nonrhyming pseudowords (O+P−) in the visual conditions (AV and VV) were plausible, making it unclear whether incorrect responses for these items reflected decoding to rhyming phonologies, strategic responses based on visual similarity or guessing. Restriction of our analyses to the 2 rhyming conditions was the most straightforward means of addressing both concerns.

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) served to localize perceptual regions for the region of interest (ROI) analyses. 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).

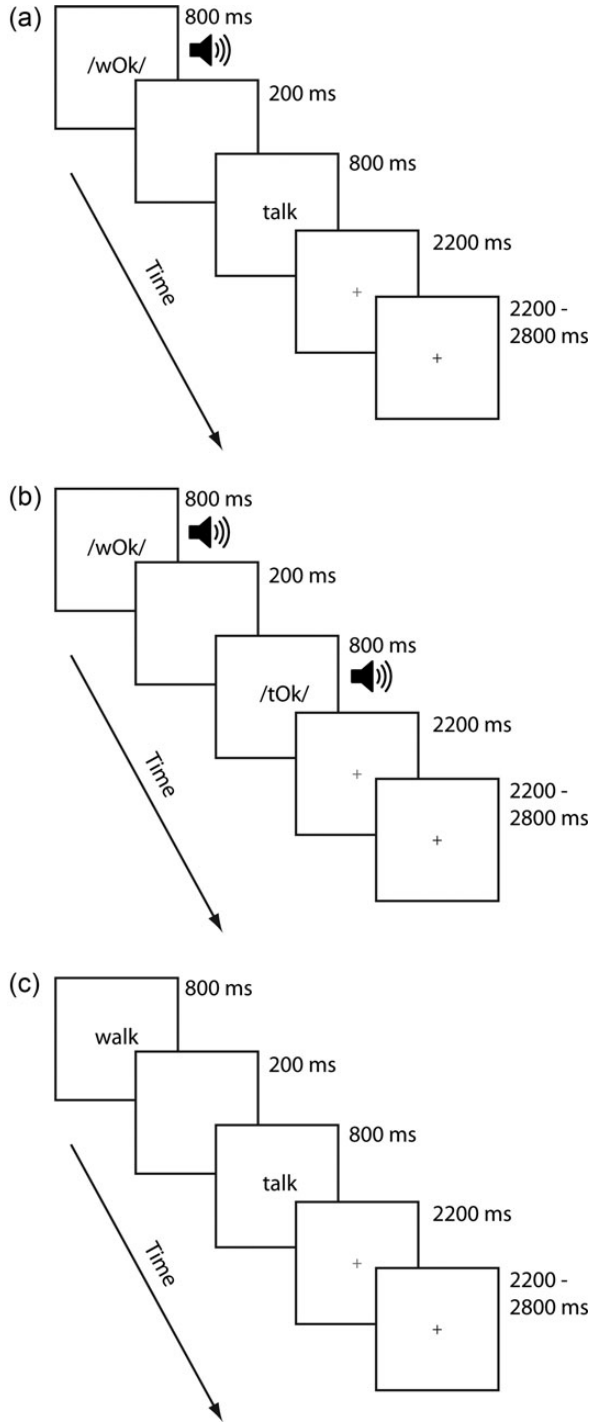

Figure 1. Schematic of trial events for the AV, AA, and VV modality conditions.

Table 1
Sample word and pseudoword stimuli used across all modality conditions

	Congruent orthography	Incongruent orthography
Rhyming		
Word	cage-rage	grade-laid
Pseudoword	punge-crunge (O+P+)	reash-sliche (O−P+)
Nonrhyming		
Word	stamp-swamp	thief-plead
Pseudoword	nouth-scouth (O+P−)	pread-slear (O−P−)

Note: Only rhyming conditions were used in the analyses.

Participants were required to determine whether the sequences matched (e.g., 2 rising sequences) or mismatched (e.g., a falling sequence followed by a steady sequence) by pressing the "yes" button to indicate a match, and the "no" button otherwise. The timing for the fixation and perceptual trials were the same as for the lexical trials.

As indicated earlier, the total duration of the 4 runs (2 word and 2 pseudoword) for each modality condition was 28 min, with additional time required to set up, speak to participants between runs and collect an anatomical reference image further increasing the total time in the scanner. Thus, to minimize fatigue, participants took part in the experiment over a number of scanning sessions within a 6-month period (mean interval between scan sessions was 5.9 weeks). During each scanning session, participants completed the rhyming task for the word and pseudoword lexicality conditions for 1 modality condition (4 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. Scheduling constraints prevented any session from including more than ~45 min of task-related scanning. 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.

MRI Data Acquisition

Participants were positioned in the MRI scanner with their head position 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. Participants wore sound attenuating headphones to minimize the effects of the ambient scanner noise. 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 T_1-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).

fMRI Preprocessing

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 replaced outlier volumes, associated with excessive movement (>4 mm in any direction) or spikes in the global signal, using interpolated values from the 2 adjacent nonoutlier scans. No more than 10% of the volumes from each run and no more than 4 consecutive volumes 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, which is an average of 152 normal adult MRI scans. This template is well defined with respect to a number of brain atlas tools and the MNI coordinate system. Moreover, stereotactic space for children within the age range included in our study has been shown to be comparable to that of adults (Burgund et al. 2002; Kang et al. 2003). Thus, it was deemed preferable to use the standard adult SPM template rather than create an average-based template, so as to compare to the previous literature. Images were smoothed using a 2 × 2 × 4 nonisotropic Gaussian kernel.

Behavioral Analyses

Because stimulus pair congruency was assumed to influence behavioral performance and BOLD activity for the task (Bitan et al. 2007),

2-way repeated-measure analysis of variance (ANOVA) tests of congruency effects were conducted with lexicality and congruency as within-subjects' independent variables. The dependent variables were the congruency effects for accuracy rates and decision latencies of correct responses.

Individual and Group-Level Image Analyses

Statistical analyses were calculated at the first level using an event-related design with all 4 lexical conditions (O+P+, O−P+, O−P−, O+P−), the fixation condition, and the perceptual condition included as conditions of interest. Interpolated volumes were deweighted, and the first 6 volumes of each run, during which a fixation cross was presented, were dropped from the analyses. A high-pass filter with a cut off of 128 s was applied. Lexical, fixation, and perceptual pairs were treated as individual events for analysis and modeled using a canonical hemodynamic response function. Voxelwise 1-sample *t*-statistic maps were generated for each participant contrasting the rhyme (O+P+ and O−P+) versus fixation conditions for each lexicality condition within each modality condition (6 contrasts), and contrasting the balanced (i.e., nonsuper-additive) cross-modal versus the unimodal conditions (AV vs. [AA + VV]) for words and pseudowords separately. Group-level results were obtained using random-effects analyses by combining subject-level summary statistics across the group as implemented in SPM8.

Region of Interest Definitions

Left auditory cortex was functionally defined within each individual by the AA > VV perceptual contrast, using a statistical threshold of $P < 0.005$ (uncorrected), masked by the aal template (Tzourio-Mazoyer et al. 2002) definition of left HG and superior temporal gyrus (STG). There were 2 clusters within this broad anatomical region that fell within the cortical atlas definitions of HG and PT, respectively. Left posterior pSTS (pSTS) was functionally defined within each individual by the union of the AA>fixation and VV>fixation condition using a statistical threshold of $P < 0.1$ (uncorrected), masked by an aal template-based mask of STS. The STS mask was the union of the aal template definitions of left middle temporal gyrus and left STG, both dilated by 4 mm along each axis. The overlapping region defines the sulcus because it follows the line that delineates these immediately adjacent atlas definitions. Posterior STS was selected by including only those voxels posterior to $y = -40$, or roughly the posterior third of the volume. A probabilistic map of the final ROI definitions across all participants is presented in Figure 2.

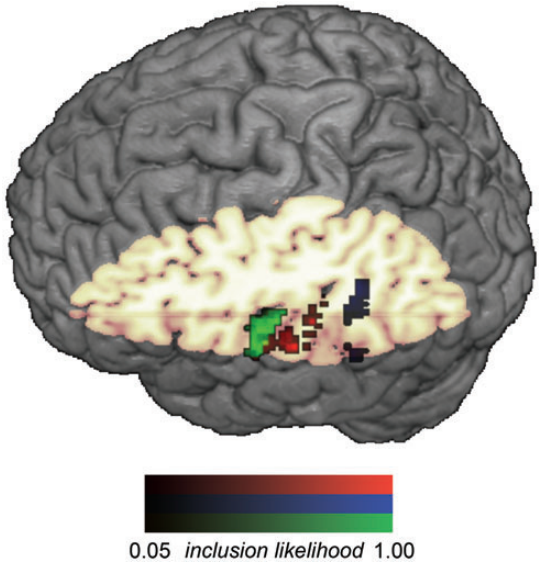

Figure 2. Probability map showing the voxelwise inclusion likelihood across participants for the intersection of unimodal (i.e., both AA and VV) congruency effect ROIs in functionally defined HG (green), PT (red). Probability map for pSTS (blue) depicts inclusion likelihood within the union of unimodal (i.e., either AA or VV) congruency effects. Voxels with the darkest values were included in the ROIs for the fewest participants, whereas voxels with the brightest values were included in the ROIs for the most participants.

Table 2
Mean (SD) decision latencies for words and pseudowords on rhyming trials

	AV	AA	VV
Words			
Congruent	1161(312)	1403 (281)	1248 (297)
Incongruent	1139 (352)	1334 (274)	1298 (306)
Pseudowords			
Congruent	1141 (429)	1372 (293)	1297 (466)
Incongruent	1186 (446)	1338 (436)	1324 (437)

Note: Decision latencies for correct responses only.

Table 3
Mean (SD) accuracy for words and pseudowords on rhyming trials

	AV	AA	VV
Words			
Congruent	0.77 (0.15)	0.77 (0.14)	0.77 (0.14)
Incongruent	0.84 (0.09)	0.84 (0.11)	0.84 (0.11)
Pseudowords			
Congruent	0.72 (0.20)	0.77 (0.16)	0.90 (0.16)
Incongruent	0.66 (0.17)	0.79 (0.15)	0.79 (0.16)

Results

Behavioral Performance

Means and SDs for decision latency and accuracy for rhyming trials in each modality condition are presented in Tables 2 and 3, and indicate that participants were accurate on the task, despite the difficulty imposed by the congruency manipulation. Two-way repeated-measure ANOVA tests of congruency effects showed a modality difference ($F_{2,42} = 8.17$, $P = 0.001$) but not a lexicality difference ($F_{1,21} = 1.18$, $P > 0.2$) with respect to decision latency, and the 2 factors did not interact ($F_{2,42} = 1.43$, $P > 0.2$). Post hoc Bonferroni-corrected pairwise comparisons indicated that the significant decision latency congruency effect difference was driven by the AA and VV conditions, with participants showing slower responses for congruent items in the AA condition and compared with the VV condition. No other decision latency differences were significant. Two-way repeated-measure ANOVA tests of congruency effects showed both modality ($F_{2,42} = 38.16$, $P < 0.001$) and lexicality differences ($F_{1,21} = 9.64$, $P < 0.001$) in the congruency effect with respect to accuracy, although the 2 factors did not interact ($F_{2,42} = 2.99$, $P > 0.05$). The modality difference was again driven by the AA and the VV conditions, with participants showing higher accuracy for congruent pairs in the VV condition compared with the AA condition. The lexicality difference was driven by greater accuracy for congruent items for pseudowords compared with words. No other differences were significant.

Both standardized literacy skill measures were correlated with the decision latency congruency effect for words in the VV condition (reading age: $r(20) = 0.76$, $P < 0.001$; spelling age: $r(20) = 0.45$, $P < 0.05$). Thus, the most skilled readers appeared to demonstrate a greater visual priming effect for congruent versus incongruent word pairs, but these effects did not extend to the auditory modality. Reading skill was not correlated with the congruency effect for accuracy.

Neuroimaging Results: Main Effects of Modality

The rhyme versus fixation contrasts defined the network of brain regions recruited for the rhyming task in each modality condition. All statistical maps were generated over the entire brain using an uncorrected $P < 0.001$, with an extent threshold calculated to obtain a cluster-level false discovery rate (FDR) corrected significance level of $q < 0.05$. Figure 3*a–c* shows the clusters reaching this significance level for the cross-modal AV condition, unimodal AA condition, and unimodal VV condition, respectively. Coordinates of peak activation within each cluster were converted to Talairach space using the mni2tal transformation function (http://imaging.mrc-cbu.cam.acuk/download/MNI2tal), and identified using the Talairach Daemon (Lancaster et al. 2000). Peak coordinates and associated statistics for these clusters are presented in Tables 4, 5, and 6.

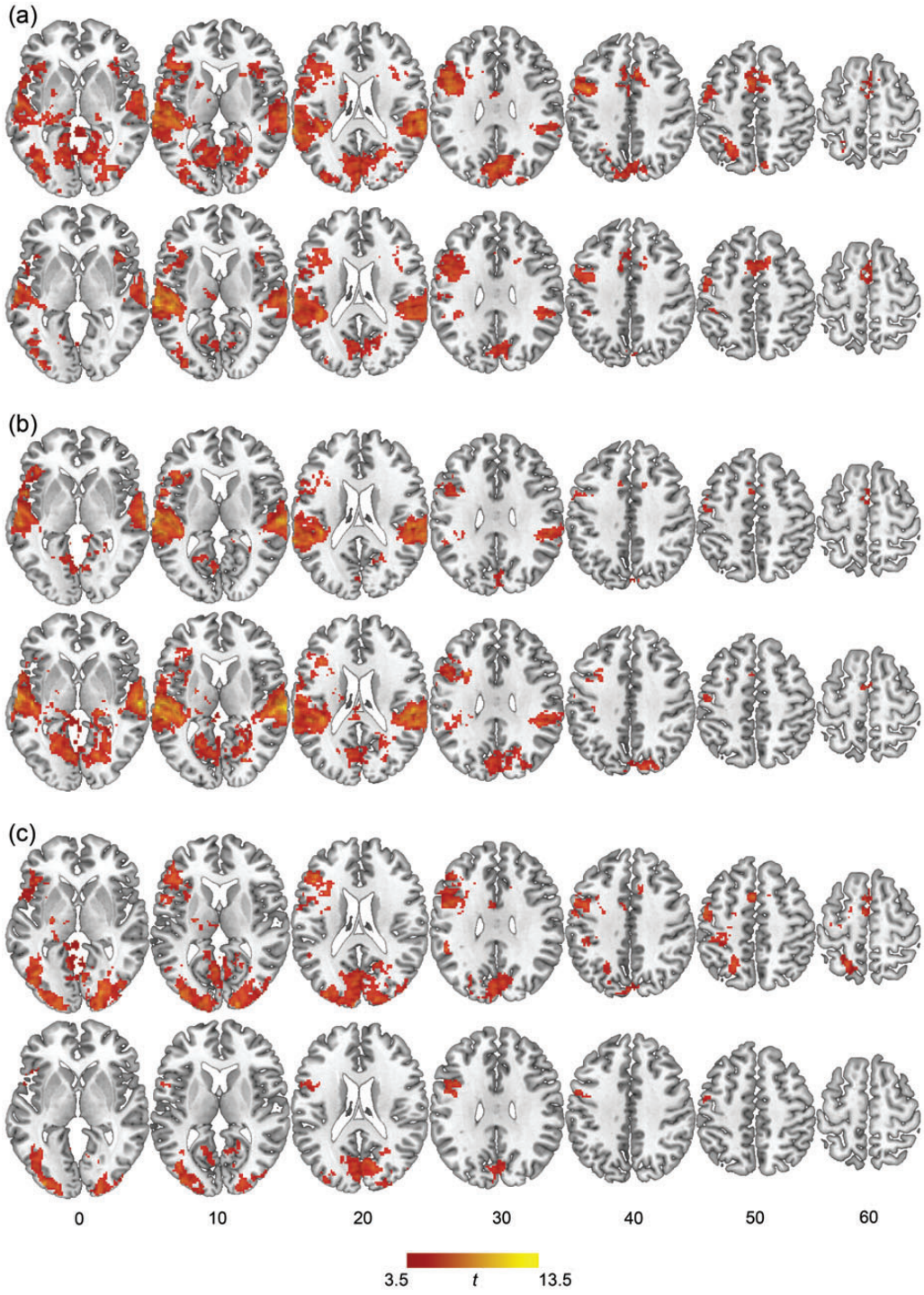

Figure 3. Whole-brain group analysis for rhyming minus fixation condition showing clusters reaching $q < 0.05$ (FDR corrected) significance in the rhyme minus fixation contrast for words (top) and pseudowords (bottom) in the AV (*a*), AA (*b*), and VV (*c*) modality conditions.

Table 4
Coordinates for clusters reaching q < 0.05 (FDR cluster-level corrected) significance in the rhyme minus fixation contrast for words and pseudowords in the AV modality condition

Region	Size	FDR	MaxZ	*x*	*y*	*z*
Words						
L Superior temporal gyrus (BA 22)	7383	<0.001	6.48	−64	−40	18
R Postcentral gyrus/insula (BA 40/13)	3141	<0.001	5.88	62	−30	18
R Precuneus/posterior cingulate (BA 19/31)	8649	<0.001	5.56	4	−74	28
L Medial frontal/superior frontal gyrus (BA 32/6)	1004	<0.001	5.31	−8	8	48
L Precuneus (BA 7)	347	<0.001	4.78	−26	−54	50
L Lentiform nucleus	150	<0.001	4.66	−18	2	6
L Caudate	100	0.002	4.55	−18	−4	22
R Claustrum	54	0.022	4.32	24	16	12
R Inferior temporal lobule (BA 20)	92	0.002	4.29	44	−8	−22
L Culmen	132	<0.001	4.25	0	−38	0
R Thalamus	53	0.022	4.22	20	−24	−2
R Cuneus (BA 19)	54	0.022	4.11	26	−88	30
R Fusiform gyrus (BA 37)	43	0.044	3.91	28	−50	−12
Pseudowords						
L Superior temporal gyrus (BA 42)	5383	<0.001	6.83	−60	−22	6
R Inferior parietal lobule (BA 40)	2281	<0.001	5.83	48	−36	26
L Superior frontal gyrus (BA 6)	859	<0.001	5	0	8	58
R Insula	347	<0.001	4.93	40	18	2
L Middle occipital gyrus (BA 19)	1098	<0.001	4.67	−38	−88	4
L Precuneus (BA 7)	1257	<0.001	4.53	−2	−74	34
R Fusiform gyrus (BA 37)	154	<0.001	4.16	38	−58	−20
L Thalamus	61	0.041	3.93	−6	−20	10

Note: L, left; R, right; BA, Brodmann area; FDR, FDR cluster-size-corrected significance level; MaxZ, peak *Z*-statistic. Size is measured in 8 mm^3voxels. Coordinates reflect standard MNI space.

Table 5
Coordinates for clusters reaching *q* < 0.05 (FDR cluster-level corrected) significance in the rhyme minus fixation contrast for words and pseudowords in the AA modality condition

Region	Size	FDR	MaxZ	*x*	*y*	*z*
Words						
L Superior temporal gyrus (BA 42)	5199	<0.001	6.64	−60	−32	12
R Superior temporal gyrus (BA 22)	2834	<0.001	6.46	60	−20	10
L Superior frontal gyrus (BA 6)	236	<0.001	4.98	−2	6	54
R Lingual gyrus (BA 17)	105	0.001	4.63	12	−70	−10
L Posterior cingulate (BA 30)	456	<0.001	4.27	−6	−72	6
R Hippocampus	40	0.034	4.2	28	−46	0
R Parahippocampal gyrus (BA 30)	74	0.004	4.18	12	−36	−6
R Lingual gyrus (BA 19)	63	0.007	4.13	12	−54	−4
R Cingulate gyrus (BA 32)	49	0.019	4.08	14	14	40
R Precuneus	80	0.003	3.98	22	−58	22
L Cuneus (BA 19)	210	0.000	3.89	−2	−90	28
L Culmen	39	0.034	3.89	−8	−44	−10
Pseudowords						
R Superior temporal gyrus (BA 22)	17 334	<0.001	6.92	62	−18	0
L Lentiform nucleus	93	0.006	4.61	−24	−6	12
L Medial frontal gyrus (BA 6)	132	0.001	4.25	−4	−2	62
L Precentral gyrus (BA 4/6)	84	0.009	4.23	−46	−14	50
L Superior temporal gyrus (BA 21)	73	0.015	4.03	−40	−8	−12
R Parahippocampal gyrus (BA 30)	122	0.002	4.02	24	−16	−16

Note: L, left; R, right; BA, Brodmann area; FDR, FDR cluster-size-corrected significance level; MaxZ, peak *Z*-statistic. Size is measured in 8 mm^3 voxels. Coordinates reflect standard MNI space.

The whole-brain random-effects analysis of the balanced cross-modal minus unimodal congruency effects (AV − [AA + VV]) did not reveal any regions in which the congruency effect for cross-modal items was reliably greater than for the unimodal conditions, except at relatively liberal uncorrected significance thresholds. This was the case whether events were modeled from the onset of the first or the second stimulus. Our key hypothesis, however, is that sensitivity of auditory cortex to the cross-modal influence of orthographic information develops with increased reading experience. Our participants were school-aged children spanning a range of

Table 6
Coordinates for clusters reaching q < 0.05 (FDR cluster-level corrected) significance in the rhyme minus fixation contrast for words and pseudowords in the VV modality condition

Region	Size	FDR	MaxZ	*x*	*y*	*z*
Words						
R Cuneus (BA 18/19)	8858	<0.001	5.8	18	−96	12
L Middle frontal gyrus (BA 46)	2977	<0.001	5.63	−44	32	24
L Medial frontal gyrus (BA 6)	405	<0.001	5.37	−2	10	52
L Superior parietal lobule (BA 7)	476	<0.001	4.94	−22	−56	56
L Inferior parietal lobule (BA 40)	363	<0.001	4.91	−50	−42	30
L Lentiform nucleus	220	<0.001	4.52	−24	−14	6
L Cingulate gyrus (BA 24/32)	81	0.015	4.49	−6	0	38
L Precentral gyrus (BA 6)	207	<0.001	4.35	−30	−16	70
R Fusiform gyrus (BA 37)	88	0.013	4.03	42	−30	−16
R Cingulate gyrus (BA 32)	79	0.015	3.9	8	12	44
L Thalamus	83	0.015	3.7	−8	−16	−8
Pseudowords						
L Inferior occipital gyrus (BA 18)	1704	<0.001	5.6	−34	−92	0
R Occipital pole (BA 18)	2132	<0.001	4.87	20	−96	6
L Middle frontal gyrus (BA 44)	587	<0.001	4.71	−40	10	32
R Lingual gyrus (BA 19)	106	0.002	4.6	18	−70	−12
L Inferior frontal gyrus (BA 47)	53	0.028	3.95	−46	20	0

L, left; R, right; BA, Brodmann area; FDR, FDR cluster-size-corrected significance level; MaxZ, peak *Z*-statistic.
Size is measured in 8 mm^3 voxels. Coordinates reflect standard MNI space.

literacy experience. Thus, our central prediction was tested in the skill-related analyses that follow.

Neuroimaging Results: ROI Analyses of Developmental Congruency Effects

Analyses were carried out in subject-specific ROIs constructed as follows: Each participant's functionally defined HG, PT, and pSTS ROI served to precisely identify these regions for each individual. Following the approach used by other multisensory integration studies (e.g., Beauchamp et al. 2004), we identified within these regions voxel populations that were sensitive to both auditory and visual congruency by taking the conjunction of the first-level congruent versus incongruent contrasts for the unimodal AA and VV conditions. The final ROIs thus identified those voxels in HG and PT that are sensitive to both unimodal auditory and visual congruency. Within pSTS, most participants did not show areas sensitive to word-level congruency for both unimodal conditions, necessitating a more inclusive definition. pSTS ROI definitions thus included voxel populations that were sensitive to congruency for "either" unimodal condition at a more liberal threshold. The mean beta values within each ROI were obtained for the congruent (O+P+) and incongruent (O−P+) lexical conditions separately for word and pseudowords in each modality condition. The difference between congruent and incongruent betas determined the congruency effect for each modality within these regions. We carried out planned correlational tests to assess whether the congruency effect for the cross-modal AV condition within these regions differs from those of the unimodal conditions as a function of literacy.

Within PT, reading age and spelling age were significantly positively correlated with the cross-modal congruency effect (reading age: $r(20) = 0.40$, $P < .05$; spelling age: $r(20) = 0.42$, $P < 0.05$) but not for either the unimodal auditory (reading age: $r(20) = -0.08$, *ns*; spelling age: $r(20) = -0.33$, *ns*) nor the unimodal visual (reading age: $r(20) = -0.09$, *ns*; spelling age: $r(20) = -0.16$, *ns*). Tests of differences between paired correlations were conducted to determine whether the correlations between cross-modal congruency and literacy measures

differed significantly from those of both unimodal conditions. These differences either approached or reached significance for reading age (AA: $t_{(20)} = 1.64$, $P = 0.05$; VV: $t_{(20)} = 1.60$, $P = 0.06$) and significantly differed for spelling age (AA: $t_{(20)} = 2.78$, $P = 0.005$; VV: $t_{(20)} = 1.93$, $P = 0.03$). Neither reading age nor spelling age were significantly correlated with the cross-modal congruency effect for pseudowords (reading age: $r(20) = -0.12$, *ns*; spelling age: $r(20) = 0.10$, *ns*). Tests of differences between paired correlations were conducted to determine whether the correlations between cross-modal congruency and literacy measures differed significantly between words and pseudowords. Cross-modal word congruency was significantly more correlated with reading age than was cross-modal pseudoword congruency, $t_{(20)} = 2.51$, $P = 0.01$, and the lexicality difference between correlations involving spelling age was marginal, $t_{(20)} = 1.47$, $P < 0.08$.

Within HG, neither the correlations between congruency effects for any modality and literacy measure, nor the differences between cross-modal and unimodal correlations were significant. Within pSTS, there were similarly no significant correlations between congruency effects for any modality and literacy measure, nor were the differences between cross-modal and unimodal correlations significant.

To summarize, literacy influenced the sensitivity of PT to cross-modal congruency, but did not influence the sensitivity of this region to unimodal congruency. Moreover, this effect was apparent only for words but not for pseudowords. Literacy did not similarly influence HG or pSTS sensitivity to word-level congruency for any modality (Fig. 4). We acknowledge a potential confound in the correlation between literacy skill and the PT congruency effect, which might be explained as resulting from strictly maturational changes across the 6 year range of participant ages. Partial correlations calculated between the PT congruency effect and each literacy measure, accounting for chronological age showed that neither reading nor spelling age were significant predictors when we accounted for the variance attributable to chronological age (reading age: $r = 0.17$; spelling age: $r = 0.31$). Conversely, neither was chronological age a significant predictor on its own when we accounted for the variance attributable to either literacy measure. This is unsurprising given the significant positive correlations between both literacy measures and chronological age (reading: $r(20) = 0.66$; spelling: $r(20) = 0.56$, both $P < 0.01$). Thus, within our sample, literacy skill is difficult to disentangle from maturational development. Note, however, that literacy was related to the cross-modal congruency effect only for real words. Were the cross-modal congruency effects driven primarily by maturational processes, one would expect this effect to be apparent in both word and pseudoword conditions. We therefore feel it most straightforward to explain these effects as arising from the increased literacy skill afforded by additional experience.

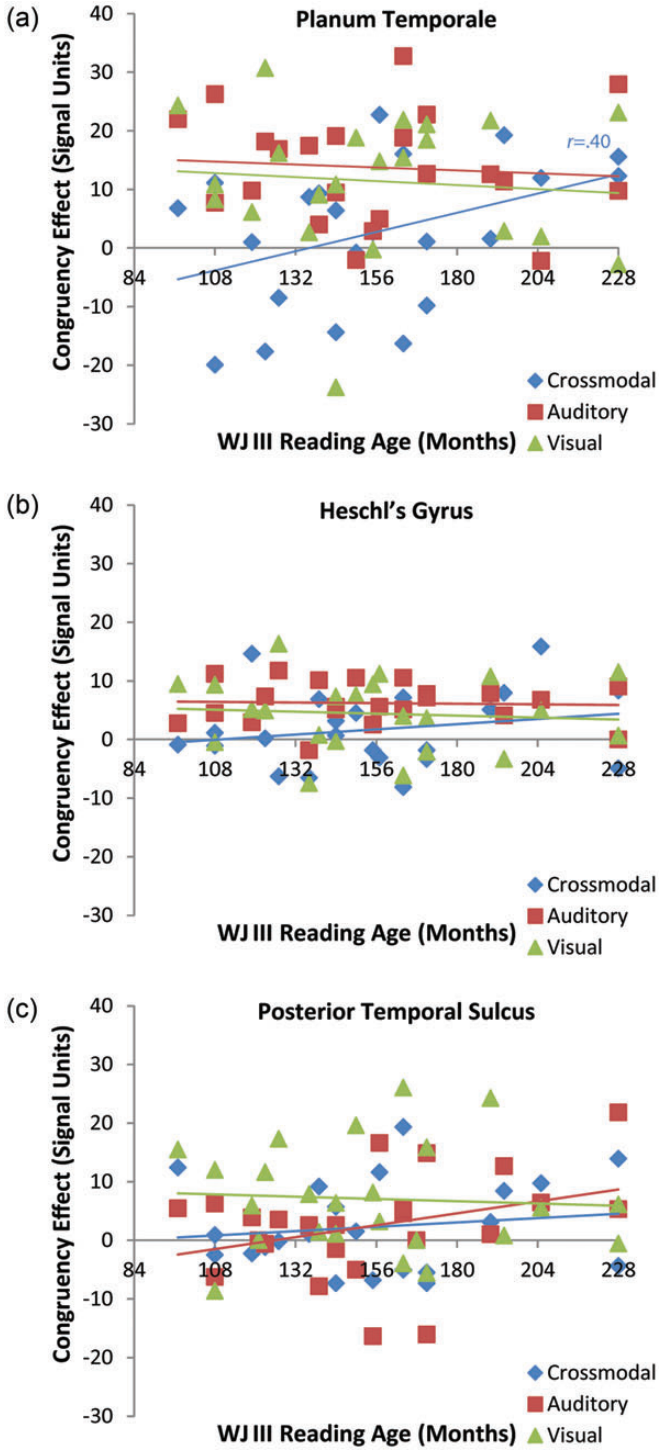

Figure 4. Literacy-related congruency effects (O+P+ minus O−P+) in planum temporale (PT) were significantly correlated with reading skill measures for the cross-modal condition but not for the unimodal conditions (*a*). No reading skill correlations were apparent in Heschl's Gyrus (*b*) or in posterior superior temporal sulcus (pSTS) (*c*). Only significant correlations are shown. The pattern of correlations with spelling skill measures was similar.

Discussion

We measured brain response to interstimulus congruency for unimodal auditory or visual and for cross-modal audiovisual presentation in auditory cortex. Although the task required decisions based on phonology, and we restricted our analyses to rhyming items (i.e., matching phonology), orthographic stimulus characteristics nonetheless had a literacy-dependent influence on processing in these regions for cross-modally presented items. These skill-related congruency and modality effects provide a strong case that literacy drives PT to become increasingly involved in multisensory integration of audiovisual information. Furthermore, that the cross-modal congruency effects were greater for words than for pseudowords,

further supports our argument that these effects arise from experience with reading and spelling words. Because the cross-modal congruency effects we observed were skill dependent, the inclusion of children with less literacy experience, for whom our analyses indicate these effects were muted, reduced overall modality differences between cross-modal and unimodal conditions when collapsed across all participants.

Literacy skill was positively correlated with behavioral congruency effects reflected in decision latencies for words in the unimodal visual condition. Literacy skill may have promoted a greater speed benefit from the orthographic overlap for 2 visually presented words, but not for pseudowords, because the orthography–phonology mappings for the latter items is unfamiliar to children of all skill levels. Although significant congruency differences were found between unimodal conditions, it should be noted that these effects were relatively small in relation to the standard deviations for these measures. Thus, the behavioral differences, although significant, were subtle, and did not include the cross-modal condition.

Cross-Modal Processing in the Dorsal Auditory Stream

Converging evidence from several methodologies has demonstrated the sensitivity of auditory cortex to cross-modal input, including animal studies employing single-cell recording (Kayser et al. 2008) and neuroimaging (Kayser et al. 2007), and in humans using functional neuroimaging (Calvert 2001; van Atteveldt, Formisano, Goebel et al. 2007). The dorsal auditory stream in particular, of which the PT is part, has been implicated in integration of auditory and articulatory-related motor information (Hickok and Saberi 2012), which unfolds over time. The dual-stream model of speech processing holds that the dorsal stream is involved in translating phonological representations into targets for the articulatory system (Hickok 2012). Although such representations are clearly useful for speech production, it has been argued that the articulatory codes generated by this stream are also important when listening to speech (Rauschecker 2011; Hickok 2012).

Although the dual-stream model posits a role of PT in processing spoken language, there has been little reliable evidence suggesting this region plays a role in reading, although methodological issues may underlie this inconsistency (Buchsbaum et al. 2005). Accordingly, much of what we know about the participation of the dorsal auditory stream in cross-modal processing comes from studies of the reception of spoken language. Our results extend this literature in several important ways: First, because the articulation of rhymes does not change as a function of orthography, the congruency effects observed in PT suggest that this region is additionally sensitive to static visual information (e.g., during reading). Second, we show that increased cross-modal lexical processing in this region is a function of literacy. Third, we investigated congruency effects arising from orthographic congruency at the rime level of sequentially presented words. That is, we manipulated associative congruency between orthographic and phonological representations, rather than temporal congruency as in previous studies that established audiovisual integration in auditory cortex within a narrow time window. Our congruency effects are grounded in the internal decoding and generation of auditory representations, rather than the temporal synchrony of external input. Thus, our results demonstrate that PT is sensitive to congruency of internally generated representations and the duration of the time window over which cross-modal integration occurs is sensitive to the nature of stimuli involved.

These results also inform models of phonological working memory of which the PT is believed to be a critical component (Buchsbaum and D'Esposito 2008). The current understanding of the phonological loop holds that spoken material is automatically encoded in the phonological store, whereas written material must first be subvocalized (Baddeley 2012). Under this assumption, the present results may be understood in the context of the dual auditory stream framework to suggest that the articulatory codes generated by PT for speech production are borrowed by processes engaged in mapping between orthography and phonology. Since it was proposed, the classical model of working memory has also been revised to allow changes within the long-term memory store to influence working memory: The most recent conceptualization, described by Baddeley (2012), explicitly provides a mechanism for characteristics of visual events, encoded in long-term memory, to influence phonological processing. The present results suggest that the applicability of the working memory model to models of reading depends on this interaction between long-term memory stores of visuo-orthographic representations and the phonological loop.

Cross-Modal Processing in the Reading Network

We interpret our findings in the context of a broader network of functionally specialized areas that have been implicated in cross-modal processing during reading. These areas include left inferior parietal lobule, an area implicated in the cross-modal mapping between orthography and phonology (Booth et al. 2002); and pSTS, a region implicated in audiovisual integration in multiple domains (Calvert 2001).

A number of studies have suggested that the pSTS plays a role in the integration of lexical stimuli (van Atteveldt et al. 2004, 2009). Preliminary whole-brain analyses of this region failed to find modality-related congruency differences within this region, and the more sensitive ROI-based approach additionally failed to find any literacy-related modality differences in pSTS. We do, however, note that the role of pSTS in multisensory integration is a complicated one. Research implicating pSTS as an audiovisual integration hub has found evidence for integration only within a relatively narrow time window. Beauchamp et al. (2010), for example, showed the McGurk effect was disrupted by transcranial magnetic stimulation of pSTS only within a 100-ms window, although van Atteveldt, Formisano, Blomert et al. (2007) found evidence of integration of phoneme-grapheme pairs in pSTS using a temporal window as large as 300 ms. Our task, in contrast, involved rhyme judgments of monosyllabic word pairs presented 1000 ms apart. Moreover, our analyses focused on between-item congruency to more closely follow previous approaches to audiovisual integration in reading carried out by van Atteveldt and co-workers. Thus, our results should not be construed to mean that pSTS plays no role in audiovisual integration for whole words, and a different experimental design

or analytic approach might reveal robust skill-related changes in cross-modal processing within this region.

From a computational perspective, our results, in relation to the body of literature showing cross-modal integration of smaller-grained lexical units within pSTS (e.g., van Atteveldt et al. 2004, 2009), suggest that PT operates on the continuous stream of audiovisual lexical objects integrated downstream (i.e., in pSTS). Under this interpretation, literacy experience drives the representations in auditory cortex become increasingly multimodal in nature, and serve as the representational targets for cross-modally mapped representations. This argument generally mirrors the functionality proposed by Golfinopoulos et al. (2010) for mapping motor signals to phonological representations in the domain of speech production. An alternative explanation is that literacy promotes the extension of the audiovisual integration functionality from pSTS into PT (Hickok 2009; Hickok et al. 2009). Our data do not distinguish between these interpretations, although we identify below several factors favoring the representational change over the functional change interpretation, at least for reading.

PT appears to be involved in representing and maintaining phonological representations, especially during reading of more phonologically complex items (e.g., word pairs) (Okada et al. 2003). The representation encoded by the current state of activity in the region resolves to the candidate representation most closely matching a stored representation (allowing for additional top-down contextual biases, such as orthographic or even semantic constraints). This account computationally parallels the process of phonological resolution of lexical stimuli during speech perception outlined in the McClelland and Elman (1986) TRACE model.

Interpreted as a skill-related representational change within PT, our results suggest that the representations active in more skilled readers take on an increasingly orthographic quality. One advantage of this change may be that the additional dimensionality afforded by adding cross-modal information to phonological representations permits the modulation of distance between representations. In computational models of orthographic to phonologic mapping (e.g., Harm and Seidenberg 1999), robust phonological representations form around attractors, which are points in multidimensional space around which representations settle. A physical analogy of an attractor in 3D space might be a large stellar body, such as a star, which exerts a strong gravitational pull on surrounding matter. As in the analog physical system, there is an inverse relationship between distance and attractive pull: Distant and close points are respectively weakly and strongly pulled toward the attractor basin. More complex representations can be potentially more distant from one another (i.e., distinct). This should benefit reading because any given representation would be less likely to be close to (and therefore influenced by) more than one competing attractor. This would have the benefit of decreasing the time required to resolve phonological representations in an orthographically deep language.

Interpreted as a functional change within PT, how cross-modal audiovisual processing capacity develops in the service of reading is less clear. Additional processing units working in parallel with pSTS might facilitate integration of more complex representations, as they permit more bits of resolution (McNorgan et al. 2011). However, because greater representational resolution should be available to congruent and incongruent stimuli alike, it is difficult to see how congruency effects might support this interpretation. Alternatively, processing from pSTS might cascade into PT for particularly difficult integration problems. Incongruent stimuli are presumably more challenging to integrate than are congruent stimuli, but this would seem to predict positive correlations between skill and incongruency-related processing (i.e., the inverse of the congruency effect), and is therefore inconsistent with our results.

We propose a framework of audiovisual integration during normal reading in which the pSTS is engaged in temporally sensitive integration, matching the incoming visual stream at smaller grain sizes (e.g., graphemes, bigrams, or trigrams) against the temporally unfolding phonological stream at correspondingly small grain sizes (e.g., phonemes). This phonological stream may be externally generated, for example, as when a child learns to match written to spoken word forms uttered by a teacher. As a reader's vocabulary expands, however, she is able to use stored phonological representations to internally generate her own phonological stream. The activation of these representations is bootstrapped from contextual cues (e.g., semantics) that suggest plausible candidates against which the orthographic stream can be matched. The product of integration at smaller grain sizes feeds forward to PT, where a larger grain-size multimodal representation emerges over time.

Our results should be considered in the context of orthographic depth, which specifies the reliability of the mappings between graphemes and corresponding phonetic representations in a particular language. In fact, our congruency manipulation was possible only because the orthographic depth of the English language permits the mapping of multiple orthographic representations to a common phonological pattern, and the converse (e.g., THREW, THROUGH, TROUGH). To reiterate our results, differential resolution of cross-modal versus unimodal congruency was apparent in auditory cortex and over asynchronously presented monosyllabic whole words. In contrast, van Atteveldt and co-workers, who have argued for a critical role of pSTS in cross-modal processing during reading, have largely done so using data from Dutch readers (e.g., van Atteveldt et al. 2004; van Atteveldt, Formisano, Blomert et al. 2007; van Atteveldt, Formisano, Goebel et al. 2007; Blau et al. 2009). The relative differences in orthographic depth of these 2 languages might explain, at least in part, the relatively inconsistent role of pSTS within this literature: Children who learn orthographically shallow languages with consistent letter-sound mappings (such as Dutch) code phonology at smaller grain sizes than do those who learn orthographically deep languages (such as English), who augment these mappings with whole-word representations (Goswami et al. 2003). These cross-linguistic differences further support an interpretation of the pSTS as an integration site at the smallest grain size, and auditory cortex as a region that maintains representations at a larger grain size in orthographically deep languages.

Conclusions

Previous studies have shown that primary and associative auditory cortex participates in the rapid automatic cross-modal integration of fine-grained orthographic and phonological representations presented within a narrow temporal

window. We extend these findings by showing that the engagement of auditory cortex in cross-modal integration appears to be experience dependent. Moreover, these effects were found using whole-word stimuli pairs presented over a much wider time window than employed in previous studies. The skill-dependent congruency effect for cross-modal stimuli observed within auditory cortex in a task reliant on the maintenance of phonological representations between successive whole-word stimulus presentations suggests that, as children become more literate, the phonological representations they maintain in memory become increasingly multimodal at the whole-word level. We note that it is difficult to disentangle the developmental changes associated with maturation and those associated with learning to read. In typically developing readers, increasing age is associated with increased literacy skill. However, it is difficult to imagine how the congruency effect might arise without literacy experience. Accordingly, it should be noted that despite the substantial variance shared between literacy skill and chronological age, chronological age was not a significant predictor of the congruency effect on its own.

These results are generally consistent with several computational models (e.g., McClelland and Elman 1986; Golfinopoulos et al. 2010) and brain-based models (van Atteveldt et al. 2009) of language processing in which auditory cortex participates in the resolution of phonological representations. We further suggest that online integration (i.e., of the perceptual stream) at smaller grain sizes in pSTS, influenced by recurrent activation from primary and associative auditory cortex, drives the formation and maintenance of increasingly multimodal (i.e., less purely phonological) representations in PT implicated as the phonological store. This finding provides insight into how acquisition of literacy impacts phonological processing by showing that experience with visual word representations changes the way fluent readers represent phonological knowledge.

Funding

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

Notes

Conflict of Interest: None declared.

References

Baddeley A. 2012. Working memory: theories, models, and controversies. Annu Rev Psychol. 63:1–29.

Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron. 41:809–823.

Beauchamp MS, Nath AR, Pasalar S. 2010. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci. 30:2414–2417.

Bitan T, Burman DD, Chou TL, Lu D, Cone NE, Cao F, Bigio JD, Booth JR. 2007. The interaction between orthographic and phonological information in children: an fMRI study. Hum Brain Mapp. 28:880–891.

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. Curr Biol. 19:503–508.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM. 2004. Development of brain mechanisms for processing orthographic and phonologic representations. J Cogn Neurosci. 16:1234–1249.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM. 2002. Functional anatomy of intra- and cross-modal lexical tasks. Neuroimage. 16:7–22.

Buchsbaum BR, D'Esposito M. 2008. The search for the phonological store: from loop to convolution. J Cogn Neurosci. 20:762–778.

Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF. 2005. Reading, hearing, and the planum temporale. Neuroimage. 24:444–454.

Burgund ED, Kang HC, Kelly JE, Buckner RL, Snyder AZ, Petersen SE, Schlaggar BL. 2002. The feasibility of a common stereotactic space for children and adults in fMRI studies of development. Neuroimage. 17:184–200.

Calvert GA. 1997. Activation of auditory cortex during silent lipreading. Science. 276:593–596.

Calvert GA. 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex. 11:1110–1123.

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. 2002. Functional neuroimaging of speech perception in infants. Science. 298: 2013–2015.

Frith U. 1985. Beneath the surface of developmental dyslexia. In: Patterson K, Marshall J, Coltheart M, editors. Surface dyslexia, neuropsychological and cognitive studies of phonological reading. London: Erlbaum. p. 301–330.

Froyen D, Bonte M, van Atteveldt N, Blomert L. 2008. The long road to automation: neurocognitive development of letter–speech sound processing. J Cogn Neurosci. 21:567–580.

Froyen D, Van Atteveldt N, Bonte M, Blomert L. 2008. Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. Neurosci Lett. 430:23–28.

Golfinopoulos E, Tourville JA, Guenther FH. 2010. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage. 52:862–874.

Goswami U, Ziegler JC, Dalton L, Schneider W. 2003. Nonword reading across orthographies: how flexible is the choice of reading units? Appl Psychol. 24:235–247.

Harm MW, Seidenberg MS. 1999. Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol Rev. 106:491–528.

Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ. 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J Neurosci. 27:7881–7887.

Hein G, Knight RT. 2008. Superior temporal sulcus—it's my area: or is it? J Cogn Neurosci. 20:2125–2136.

Hickok G. 2012. The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. J Commun Disord. 45:393–402.

Hickok G. 2009. The functional neuroanatomy of language. Phys Life Rev. 6:121–143.

Hickok G, Okada K, Serences JT. 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 101:2725–2732.

Hickok G, Saberi K. 2012. Redefining the functional organization of the planum temporale region: space, objects, and sensory–motor integration. In: Poeppel D, Overath T, Popper AN, Fay RR, editors. The human auditory cortex. New York (NY): Springer. p. 333–350.

Kang HC, Burgund ED, Lugar HM, Petersen SE, Schlaggar BL. 2003. Comparison of functional activation foci in children and adults using a common stereotactic space. Neuroimage. 19:16–28.

Kayser C, Petkov CI, Augath M, Logothetis NK. 2007. Functional imaging reveals visual modulation of specific fields in auditory cortex. J Neurosci. 27:1824–1835.

Kayser C, Petkov CI, Logothetis NK. 2008. Visual modulation of neurons in auditory cortex. Cereb Cortex. 18:1560–1574.
Koelewijn T, Bronkhorst A, Theeuwes J. 2010. Attention and the multiple stages of multisensory integration: a review of audiovisual studies. Acta Psychol (Amst). 134:372.
Kuhl PK. 2004. Early language acquisition: cracking the speech code. Nat Rev Neurosci. 5:831–843.
Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. 2000. Automated Talairach Atlas labels for functional brain mapping. Hum Brain Mapp. 10:120–131.
Liberman AM, Mattingly IG. 1985. The motor theory of speech perception revised. Cognition. 21:1–36.
McClelland JL, Elman JL. 1986. The TRACE model of speech perception. Cognit Psychol. 18:1–86.
McGurk H, Macdonald J. 1976. Hearing lips and seeing voices. Nature. 264:746–748.
McNorgan C, Alvarez A, Bhullar A, Gayda J, Booth JR. 2011. Prediction of reading skill several years later depends on age and brain region: implications for developmental models of reading. J Neurosci. 31:9641–9648.
Miller LM, D'Esposito M. 2005. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci. 25:5884–5893.
Nath AR, Beauchamp MS. 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage. 59:781–787.
Nath AR, Fava EE, Beauchamp MS. 2011. Neural correlates of interindividual differences in children's audiovisual speech perception. J Neurosci. 31:13963–13971.
Okada K, Smith KR, Humphries C, Hickok G. 2003. Word length modulates neural activity in auditory cortex during covert object naming. Neuroreport. 14:2323–2326.
Rauschecker JP. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res. 271:16–25.
Ross LA, Molholm S, Blanco D, Gomez-Ramirez M, Saint-Amour D, Foxe JJ. 2011. The development of multisensory speech perception continues into the late childhood years. Eur J Neurosci. 33:2329–2337.
Schlaggar BL, McCandliss BD. 2007. Development of neural systems for reading. Annu Rev Neurosci. 30:475–503.
Shaywitz SE, Shaywitz BA, Fletcher JM, Escobar MD. 1990. Prevalence of reading disability in boys and girls: results of the connecticut longitudinal study. JAMA. 264:998–1002.
Stevenson RA, James TW. 2009. Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. Neuroimage. 44:1210–1223.
Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273–289.
van Atteveldt N, Formisano E, Blomert L, Goebel R. 2007. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex. 17:962–974.
van Atteveldt N, Formisano E, Goebel R, Blomert L. 2004. Integration of letters and speech sounds in the human brain. Neuron. 43:271–282.
van Atteveldt N, Roebroeck A, Goebel R. 2009. Interaction of speech and script in human auditory cortex: insights from neuro-imaging and effective connectivity. Hear Res. 258:152–164.
van Atteveldt NM, Formisano E, Goebel R, Blomert L. 2007. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. Neuroimage. 36:1345–1360.
Wechsler D. 1999. Wechsler abbreviated scale of intelligence. San Antonio (TX): The Psychological Corporation.
Woodcock RW, McGrew KS, Mather N. 2001. Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing.
Zeno S. 1995. The educator's word frequency guide. Brewster (NY): Touchstone Applied Science Associates.
Ziegler JC, Goswami U. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. Psychol Bull. 131:3–29.