Cerebral Cortex March 2010;20:622-632
doi:10.1093/cercor/bhp129
Advance Access publication July 1, 2009

Auditory Selective Attention to Speech Modulates Activity in the Visual Word Form Area

Yuliya N. Yoncheva[1], Jason D. Zevin[1], Urs Maurer[1,2] and Bruce D. McCandliss[1,3]

[1]Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York, NY, USA, [2]Department of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland and [3]Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

Selective attention to speech versus nonspeech signals in complex auditory input could produce top-down modulation of cortical regions previously linked to perception of spoken, and even visual, words. To isolate such top-down attentional effects, we contrasted 2 equally challenging active listening tasks, performed on the same complex auditory stimuli (words overlaid with a series of 3 tones). Instructions required selectively attending to either the speech signals (in service of rhyme judgment) or the melodic signals (tone-triplet matching). Selective attention to speech, relative to attention to melody, was associated with blood oxygenation level–dependent (BOLD) increases during functional magnetic resonance imaging (fMRI) in left inferior frontal gyrus, temporal regions, and the visual word form area (VWFA). Further investigation of the activity in visual regions revealed overall deactivation relative to baseline rest for both attention conditions. Topographic analysis demonstrated that while attending to melody drove deactivation equivalently across all fusiform regions of interest examined, attending to speech produced a regionally specific modulation: deactivation of all fusiform regions, except the VWFA. Results indicate that selective attention to speech can topographically tune extrastriate cortex, leading to increased activity in VWFA relative to surrounding regions, in line with the well-established connectivity between areas related to spoken and visual word perception in skilled readers.

Keywords: complex sounds, fusiform gyrus, pure-tone judgment, rhyming, speech perception

Introduction

Speech perception often occurs in a densely cluttered, rapidly changing acoustic environment, where multiple sounds vie for attention. Thus, successful communication relies on focusing selectively on the relevant auditory attributes while filtering out the irrelevant inputs. Despite the importance of such cognitive processes in ecologically valid settings, the role of attention is often overlooked in investigations and theories of speech perception. It has been generally established that focusing attention on a particular input modality, a spatial location, or a given set of target features modulates cortical activity such that task-relevant representations are enhanced at the expense of irrelevant ones (e.g., Hillyard et al. 1973; Haxby et al. 1994; Luck and Hillyard 1995; Laurienti et al. 2002; Foxe and Simpson 2005). The importance of bottom-up (input-driven) and top-down (schema-driven) attentional interactions for complex auditory scene analysis has been well-documented outside the realm of speech processing (Bregman 1990). The early sensory mechanisms at play when directing auditory attention based on spatial and nonspatial cues have also been mapped (recently reviewed in Fritz et al. 2007). Yet, the role of top-down attention in shaping cortical responses specifically during speech perception remains to be isolated.

Many investigations have focused on manipulations of bottom-up stimulus properties in order to dissociate cortical areas critical for the processing of speech versus well-controlled nonspeech sounds (e.g., Binder et al. 2000). Studies of this kind have commonly linked speech perception functions to activations in temporal cortical regions, such as superior and middle temporal gyri, as well as inferior frontal areas (for review, see Demonet et al. 2005). However, differences between speech and nonspeech signals with respect to particular acoustic properties and complexity can result in stimulus-driven effects that will be intricately confounded with contrasts between linguistic and nonlinguistic processes, as these may, or may not, engage top-down mechanisms in the same fashion. Therefore, inferring specific associations between functional regions in the brain and top-down attentional processes requires experimentally manipulating the form of processing that is voluntarily carried out on identical stimuli.

One elegant solution to fully equating acoustic variation capitalizes on the fact that synthetic sine-wave syllable analogues are typically perceived as nonspeech at first, but after sufficient exposure or debriefing, come to be perceived as intelligible speech (Dehaene-Lambertz et al. 2005). Neuroimaging investigations using such sine-wave analogues have recently provided evidence for distinct cortical responses in left posterior temporal regions when experiencing the same stimuli initially as nonspeech versus subsequently as speech (Dehaene-Lambertz et al. 2005; Dufor et al. 2007). Although valuable, this experimental paradigm poses certain limitations to elucidating the role of top-down attention during speech perception. First, the intrinsic salience and semantic interpretability characteristic of naturally produced words are largely discounted in sine-wave syllable analogues. Moreover, the simple discrimination tasks employed so far do not explicitly control processing demands, thus task difficulty and degree of top-down linguistic focus might vary drastically between the 2 conditions, which, in turn, would affect the profile of cortical responses before versus after the subjects' switch to speech mode. Finally, the inherently unidirectional nature of the debriefing procedure (from nonspeech to speech) restricts this approach in its utility as a tool for investigating top-down attention to linguistic content via repeated, within-subject measures using functional magnetic resonance imaging (fMRI).

The central aim of the present study is to differentiate the cortical effects of top-down attention to linguistic versus equally challenging nonlinguistic aspects of auditory input. Attentional processes are best manifested and investigated in the presence of conflict and the need for selection (Desimone

and Duncan 1995). Here these demands are increased by presenting complex chimeric auditory stimuli that consist of auditory words overlaid with tone triplets, under task conditions necessitating selective auditory attention to 1 of the 2 dimensions, while disregarding the other. The chimeric nature of the stimuli allows holding constant bottom-up stimulus properties, while contrasting 2 active listening processing goals (rhyming versus tone-triplet judgment task) that focus attention on linguistic versus nonlinguistic (melodic) content, respectively. Linguistic processing in the current study is probed via demanding rhyming judgments that require attention to relatively fine-grained phonetic contrasts in the presence of acoustically similar distractors. According to some theorists this type of attention to segmental detail may tap orthographic knowledge in addition to, or in lieu of, the more holistic processing typical of normal speech perception (Faber 1992; Port 2007). One striking piece of evidence supporting this view comes from studies of illiterate adults who have no observable difficulty with verbal communication yet are grossly impaired at tasks that require treating speech sounds as individual segments (Morais et al. 1986).

Thus, an additional aim of this study is to examine how selective attentional focus on the phonological aspects of auditory words instantiated by these rhyming judgments would affect activity of extrastriate regions, which are typically engaged in visual processing of written words. It has often been proposed that in the process of acquiring literacy (Bradley and Bryant 1983) representations related to the visual and spoken word forms come to influence one another in a form of interactive activation (McClelland and Rumelhart 1981; Seidenberg and McClelland 1989; Grainger and Ferrand 1996). In support of this notion, auditory rhyming judgment experiments (Seidenberg and Tanenhaus 1979) have shown behaviorally that arriving at a decision that 2 auditory words rhyme is faster when the pairs are orthographically similar (e.g., *pie-tie*) than when they are orthographically dissimilar (e.g., *rye-tie*); conversely, rejecting nonrhyming auditory pairs that have overlapping spelling patterns (e.g., *couch-touch*) increases response latencies. Notably, in both cases no visual print is presented and spelling information does little to benefit rhyming judgment performance because consideration of spelling would just as likely lead to the correct response as to the incorrect response.

More recently, neuroimaging investigations have identified regions of the extrastriate visual system linked to processing orthographic aspects of visual word forms. In particular, a region in the left mid-fusiform gyrus (FG) has been termed the visual word form area (VWFA) (McCandliss et al. 2003a; Cohen and Dehaene 2004) owing to its important role in bottom-up perceptual encoding of orthographic properties of letter strings. Passive presentation of auditory words typically does not recruit this area (Cohen et al. 2004). Linguistic processing demands, however, might modulate VWFA activity in a top-down fashion (Demonet et al. 1994; Booth et al. 2002; Bitan et al. 2005) pointing to an integrative function of the VWFA as an interface between visual word form features and additional representations associated with auditory words (Schlaggar and McCandliss 2007). Differential cortical activity in left occipito-temporal regions during demanding auditory linguistic tasks has been previously reported (Demonet et al. 1994; Booth et al. 2002; Bitan et al. 2005; Cone et al. 2008). Notably, since these studies were not focused on investigating the role of selective attention, the experimental design and contrasts were not aimed at ruling out the possibility that the observed effects were associated with differences in bottom-up stimulation or general processing difficulty (as operationalized by performance measures), both known to affect activity in the VWFA (Booth et al. 2003; Binder et al. 2006).

The present study seeks to isolate top-down focus on linguistic aspects of auditory words by controlling for confounding factors (stimulation type and differences in the overall level of attentional demands) in order to test the hypothesis that selective auditory attention to language modulates responses in cortical regions involved in speech processing. We also hypothesize that top-down activation of phonological aspects of auditory words and their associated orthographic visual representations modulates BOLD activity in the VWFA. One approach to examining such effects is regarding the VWFA in isolation, considering only how activation *within* this region is modulated relative to a control condition, independent of the activity levels in surrounding extrastriate regions. Alternatively, as employed in the current paper, the top-down attentional effect could be investigated *across* regions assessing VWFA activity *relative* to neighboring extrastriate activity in a topographic fashion (Haxby et al. 1994).

Methods

Participants

Twelve healthy, right-handed, native English-speaking volunteers (mean age: 27.2 years, range: 24.8–30.2; 5 women) took part in the study. All subjects had normal vision, hearing, and reading abilities (Age-based Relative Proficiency Index for Basic Reading Skill cluster: average 98/90, minimum 96/90; Woodcock et al. 2001). All participants were fully briefed and provided written informed consent. Ethical approval was granted by the Institutional Review Board of the Weill Medical College of Cornell University. All experiments were conducted in accordance with the guidelines of the Code of Ethics of the World Medical Association (Declaration of Helsinki; 18 July 1964).

Stimuli

An auditory word (mean duration = 479 ms, SD = 63) was simultaneously presented with a tone triplet (total duration 475 ms) to form a chimeric word/tone stimulus (Fig. 1). Stimulus presentation was controlled by E-prime software (Psychology Software Tools, Inc., Pittsburgh, PA).

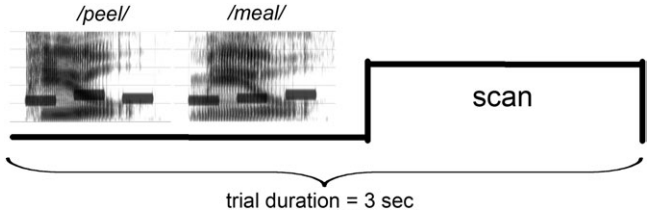

Figure 1. Schematic diagram of an active task trial, including spectrograms of 2 example stimuli. Each chimeric auditory stimulus (mean duration = 475 ms) consisted of a spoken English word presented simultaneously with a tone-triplet (a series of 3 pure tones, see the 3 horizontal bars in each spectrogram). One such stimulus was followed shortly (100 ms) by a second stimulus. Based on the preceding instructions, participants performed either a rhyming judgment task on the word pair (e.g., */peel/* and */meal/* rhyme), or a tone-triplet matching task on the pair of tone-triplets (e.g., the tone-triplets are not identical). After the presentation of the stimulus pair (total duration = 1200 ms, silent gaps pre- and poststimulus ~total 540 ms), a functional scan was acquired (clustered acquisition time ~1260 ms). Example stimuli in mp3 format can be accessed online in Supplementary Materials.

Auditory Words

A set of 256 different auditory words, each belonging to 1 of 32 rhyme "families" (example of a rhyme family: lane, crane, stain, train) was compiled. Each word was presented twice over the course of the experiment: once as a member of rhyming word pair and once as a member of a nonrhyming word pair. No heterographic homophones were included in the experimental lists, thus each auditory word was associated with a unique spelling. Two independent native English-speaking raters listened to the auditory stimuli while transcribing each word. Exact spelling match accuracy for the entire set of experimental stimuli ranged from 96.9 to 97.7%. Participants in the fMRI study heard stimuli from half of the rhyming families in the context of the rhyme focus condition and the other half in the melodic focus condition (counterbalanced across subjects).

Tones

A sequence of 3 unique pure tones constituted a tone-triplet. Pure tones corresponded to D, E, F#, G, A, B, or C# on the D major equal-tempered scale, and ranged in pitch from 1174.66 to 2217.46 Hz.

Procedure

Prior to the scanning session participants practiced the melodic focus task on a separate set of chimeric word/tone stimuli in a staircase test that progressively reduced tone amplitude while holding word amplitude constant. The sound amplitude level at which a subject reached an accuracy threshold of 90% on 2 consecutive 10-trial sessions was set as the stimulus presentation level during scanning.

fMRI Tasks

In the scanner, 2 tasks were performed on the same auditory chimeric word/tone stimuli as a 2-alternative forced choice decision: 1) in the rhyme focus condition participants judged whether the words in the stimulus pair rhymed; and 2) in the melodic focus condition whether the tone-triplet pairs were the same or not. In order to maximize the need for intensive phonological processing in the rhyme focus condition, nonrhyming trials were comprised of close distractors (distractors that shared either identical vowels and ended in phonologically similar consonants, or shared phonologically similar vowels and ended in identical consonants, e.g., blaze vs. noise). In order to promote intensive melodic analysis in the melodic focus condition, nonmatching tone-triplets were constructed by reversing the order of the second and third tones of the triplet. To ensure that rhyming decisions were based on acoustic/phonological attributes rather than spelling associations, half of all rhyme targets and distractors shared spellings of rhymes, whereas the other half did not (Seidenberg and Tanenhaus 1979). Eight runs (4 rhyme focus and 4 melodic focus tasks, alternating) were completed in the scanner. A run consisted of 9 blocks (each block lasting 24 s): 4 active blocks of the same active task, alternating with 5 fixation "rest" blocks (the first block in a run being "rest"). Each active block contained 8 trials. Each trial lasted 3 s in the following sequence: 190 ms silence, 1200 ms on average of auditory stimulus pair presentation (first and second stimuli in a pair were separated by a fixed 100 ms silent gap), 350 ms silence, and 1260 ms clustered image acquisition.

Memory Test

Following the functional scans, participants were presented with a surprise word visual recognition test to assess the relative influence of the 2 focus conditions on memories for phonological rhyme information. Twenty-five target words were chosen from the rhyme focus condition and 25 words were chosen from the tone focus condition. Each attentional focus condition involved multiple words selected from a rhyme family uniquely assigned to that condition (counterbalanced across subjects), thus allowing the matching of each target to a novel distractor item that shared this unique rhyme information. This resulted in 25 novel distractors from the rhyme family presented in the rhyme focus condition and 25 novel distractors from the rhyme family presented in the melodic focus condition. Note that matching distractor items for rhyme-level information that was presented multiple times within each attentional focus condition likely increases the subjective familiarity of these distractor items, as it presents conditions known to induce false memories (Deese 1959; Sommers and Lewis 1999). Thus, although the design of this memory test provides a potentially sensitive assay of differential processing of phonological rhyme information during the rhyme focus versus the melodic focus conditions, the choice of the distractor items (containing repeatedly presented rhyme information) likely diminishes this assay's sensitivity at the item-specific word level.

Data Acquisition

Functional (and structural) magnetic resonance imaging was performed with a GE 3 Tesla scanner equipped with an 8-channel head coil. High-resolution, T_1-weighted anatomical reference images were obtained using a 3D magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence. Functional T_2^*-weighted imaging used a spiral in-out sequence (Glover and Law 2001) with the following parameters: TR = 3 s, TE = 30 ms, flip angle = 90°, FOV = 22 cm, matrix = 64 × 64, 5 mm slice thickness, gap = 1 mm. Using a clustered acquisition protocol allowed stimulus presentation in the quiet gaps (TA = 1.26 s) when the acoustic scanner noise was absent. Sixteen oblique slices (anterior commissure/posterior commissure aligned) were acquired per volume, fully covering occipital and temporal cortices in each participant, with a maximal superior extent of the group average of $z = 30$. Each functional run lasted 228 s during which 76 volumes were collected.

Data Analysis

Behavioral measures: reaction times (RTs) for correct trials (reported in ms from the onset of the second stimulus of a pair) and accuracy (% correct responses) were analyzed to assess processing difficulty in the 2 attentional focus conditions. RTs greater than 2 SDs from the mean (M) for each task in each subject were excluded to minimize the influence of outliers.

Imaging data were analyzed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/software) in 3 major stages: preprocessing to retrieve functional data and map subjects into a common stereo-tactic space, whole-brain statistical parametric mapping, and region of interest (ROI) analyses. After discarding the first 4 images of each session to allow for T_1 equilibrium, slice-timing correction was applied to account for the fact that slices were acquired in a fixed order during the 1.26 s TA in each 3 s TR. Next, to correct for subtle head movements, image realignment was performed, generating a set of realignment parameters for each run and a mean functional image, which was used to coregister functional scans to participant's structural scan. Finally, images were normalized to the Montreal Neurological Institute (MNI) 152-mean brain, and smoothed with a 9-mm full-width half-maximum isotropic Gaussian filter, followed by re-sampling into isometric 2×2×2 mm voxels. A 2-level statistical analysis approach was applied. Correct and incorrect trials from each focus condition were modeled separately. Reported results are based on correct trials only. Condition effects in each participant were estimated using a general linear model after convolving the onset of each trial type with a canonical hemodynamic function, including the realignment parameters as covariates. Statistical parametric maps were computed for each contrast of interest (the correct trials for each condition), and these contrast maps were entered into a second-level model treating individual subjects as a random effect. Statistical significance threshold in the whole-brain analysis was set to false discovery rate (Fdr) corrected $P < 0.05$. Considering our *a priori* interest in modulations of the VWFA, in light of the exploratory finding of extensive deactivations across extrastriate cortex, we next conducted ROI analyses of FG to assess the relationship between the attentional modulation and regional patterns of deactivation.

ROI Selection and Regional Quantification of FG Activity

We performed a regional analysis of FG to further investigate the pattern of extrastriate deactivation and to directly test whether differential deactivations across fusiform regions reflect regionally specific effects as opposed to mere thresholding differences. Motivated by findings suggesting functionally significant anterior–posterior gradients in FG that differ across hemispheres (Vinckier et al. 2007;

Brem et al. 2009), and following the report of differential activity when subdividing the FG into non-overlapping anterior, middle, and posterior ROIs (Xue and Poldrack 2007), we constructed a matrix of fusiform ROIs defined based on anatomical considerations. First, the FG was divided into 3 portions that spanned equidistantly along the anterior–posterior axis. Using the population-based probabilistic maps provided by the SPM Anatomy toolbox v1.6 (Eickhoff et al. 2005, 2007), the anterior–posterior extent of the FG was identified based on the points producing 0% probability of designation to neighboring regions (i.e., $y = -32$ to $y = -86$). The automated anatomical labeling (AAL) left and right FG templates (Tzourio-Mazoyer et al. 2002) were then separately subjected to conjunction with 1 of 3 boxes (each spanning equidistantly on the $y = -32$ to $y = -86$ extent while covering the fusiform range on the x- and z-axes), thus subdividing it into an anterior, a middle, and a posterior fusiform portion. For each of these portions the center of mass was computed providing a center for an ROI: left anterior (–31.0, –41.7, –18.2), left middle (–34.7, –58.2, –14.8), left posterior (–30.4, –75.2, –13.6); right anterior (34.4, –41.2, –18.1), right middle (33.2, –58.4, –14.3), right posterior (30.4, –74.6, –12.4). Next, we created identically sized spherical ROIs. The radius of the spheres was established empirically as 8 mm: the value that provided the maximally sized non-overlapping spheres for each of the six ROIs. This resulting anatomical segregation is in general agreement with the fusiform coverage and functional distinctions suggested by reports of activations across the fusiform visual word form system using different stimulus characteristics and paradigms (Vinckier et al. 2007; Xue and Poldrack 2007; Brem et al. 2009). MarsBar (Brett et al. 2002) was used to extract data from the voxels specified within each ROI in the form of average percent signal change from rest for each active condition (i.e., rhyme focus, melodic focus) for each subject. All reported coordinates are in MNI stereotactic space.

Results

Behavioral Performance

No significant behavioral differences between the rhyme and melodic focus conditions were present based on in-scanner reaction time and accuracy measures. Both accuracy (rhyme: M = 85.10%, SD = 7.62 vs. melodic focus: M = 89.16%, SD = 6.33: $t_{11} = 1.76$, $P = 0.21$) and reaction times (rhyme: M = 880.16 ms, SD = 131.51 vs. melodic focus: M = 857.24 ms, SD = 120.07: $t_{11} = 1.56$, $P = 0.24$) were comparable between the 2 conditions.

Additional behavioral analyses were conducted to investigate the extent to which to-be-attended versus to-be-ignored stimulus information influenced decision making. These analyses contrasted performance on trials in which the content to be ignored led to a congruent response (i.e., rhyming and tone-triplet judgment led to the same response) with trials leading to an incongruent (opposite) response. Within each focus condition, a *t*-test revealed no significant effect of congruency on accuracy (rhyming task: $t_{11} = 0.002$, $P = 0.97$, tone task: $t_{11} = 0.01$, $P = 0.91$) or reaction times (rhyming task: $t_{11} = 0.03$, $P = 0.86$, tone task: $t_{11} = 0.05$, $P = 0.82$).

An analysis of response latencies in the rhyme focus condition examined the potential interaction between rhyming/nonrhyming words pairs and congruent/incongruent associated word spellings originally reported by Seidenberg and Tanenhaus (1979). A 2 × 2 ANOVA of reaction times with factors rhyming (rhyming, nonrhyming) and spelling (congruent, incongruent) did indeed reveal an interaction in the predicted direction, with relatively faster responses for rhyme trials sharing spellings and relatively slower responses for nonrhyme trials with similar spellings, but this effect fell short of significance ($F_{1,11} = 2.92$, $P = 0.12$).

Finally, performance on the post-scan surprise memory test was used to assess whether participants attended to phonological rhyme information more under the rhyme focus task than under the melodic focus task. When recognition responses were analyzed at the level of word-specific information, by contrasting target words presented in the scan with distractor words selected from the same rhyme families, in line with previous work on phonological false memories (e.g., Sommers and Lewis 1999), no significant differences appeared for items associated with either attention condition (rhyme focus $d' = 0.24$ vs. melodic focus $d' = -0.04$: $t_{11} = 1.55$, $P = 0.15$). However, when responses were analyzed at the level of phonological rhyme information, differential results emerged across the 2 attention conditions. Because nonoverlapping rhyme families were assigned to each of the attention conditions, and distractor items were selected from these segregated rhyme families, it was possible to collapse over target and distractor items to examine whether memory test items were more likely to be endorsed as "recognized" when they shared the phonological rhyme information presented under one attention condition versus the other. Memory test items selected from rhyme families assigned to the rhyme focus condition were more likely to be scored as recognized than corresponding memory test items selected from rhyme families assigned to the melodic focus condition ($t_{11} = 5.36$, $P < 0.0005$). Across targets and distractors combined, items that shared phonological rhyme information with words presented in the rhyme focus condition accounted for 65.0% of all endorsements versus 35.0% for the melodic focus condition. Further, specifically examining erroneous endorsements of distractor items also revealed a significant effect of attention condition ($t_{11} = 3.70$, $P < 0.005$). Distractor words selected from rhyme families assigned to the rhyme focus condition accounted for 61.3% of all erroneous endorsements versus 38.7% for the melodic focus condition. These 2 sets of results support the claim that phonological rhyme information was processed to a greater extent in the rhyme focus condition than in the melodic focus condition.

fMRI Results

First, via whole-brain analysis we examined BOLD responses during the active task (collapsed across condition) relative to rest. We then identified regions that were differentially active in the rhyme versus melodic focus. Second, to characterize the impact of the 2 focus conditions on activity in extrastriate visual regions, ROI-based topographical analyses of percent signal change (for each active condition versus rest) were conducted across 6 anatomically defined ROIs in the FG. This test provided an analysis of each active condition versus rest within every fusiform ROI as well as a topographical analysis of *relative* signal change *between* different ROIs.

Whole-Brain Analysis. We first examined the pattern of BOLD responses collapsed over the 2 focus conditions to establish that the densely clustered acquisition protocol successfully activated auditory regions. Results generally replicated previous findings of extensive activations in temporal cortices (Zevin and McCandliss 2005) and are displayed in Table 1 (active task > rest). Deactivations were also observed (Table 1: rest > active task) with the largest clusters spanning posterior medial regions (e.g., occipital regions, precuneus, and cuneus) and anterior medial regions (e.g., middle orbito-frontal areas).

Such task-independent BOLD decreases in these regions are typically associated with the *default network* (reviewed in Gusnard et al. 2001). The third most prominent cluster of deactivation included the most anterior portion of FG, a finding consistent with reports of modality-specific BOLD decreases during demanding auditory tasks (McKiernan et al. 2003).

Task effects: rhyme versus melodic focus conditions. The rhyme > melodic focus comparison revealed left-lateralized activations in inferior frontal gyrus (IFG)/sulcus, FG, middle temporal gyrus (MTG), and inferior temporal sulcus (ITS), as well as in right superior temporal gyrus (STG)/sulcus (STS) and right cerebellum (Table 2, Fig. 2). The left IFG cluster encompassed pars triangularis, pars orbitalis, pars opercularis, and anterior insula. The left FG cluster fell within the boundaries of the region commonly referred to as the VWFA (McCandliss et al. 2003a). The melodic > rhyme focus condition contrast produced no significant activations (even at a liberal threshold of Fdr-corrected $P < 0.1$).

ROI Topographic Analyses in Fusiform Regions

As reported above, extensive deactivations (rest > active task; Fig. 3, Table 1) manifested throughout extrastriate cortex. This deactivation effect was quantified separately for each focus condition in each of the 6 FG ROIs designed to segregate posterior, mid-, and anterior fusiform regions within each hemisphere (Table 3).

To examine whether the topographic distribution of the top-down modulation differed significantly across tasks, we quantified percent signal change between baseline rest and

Table 1

Active task (rhyme and melodic focus) versus baseline rest contrasts

MNI coordinates			Anatomical location			Statistical values		
			Peak voxel	Nearest region for this volume				
x	*y*	*z*	Location	Location	Distance (mm)	*N* voxels	*Z* voxel	$P_{Fdr\text{-}corr}$
Active task > Rest								
−54	−36	6	L MTG	L STS	4.00	6049	6.03	0.000*
−56	−24	4	L STS	L MTG	4.00		5.50	0.000*
−68	−28	6	L MTG	L STS	2.00		5.11	0.000*
54	−44	6	R MTG	R STS	5.66	4544	5.36	0.000*
64	−36	8	R STS	R MTG	2.00		5.14	0.000*
66	−14	−8	R STS	R MTG	2.00		5.02	0.000*
28	−64	−32	R Cerebellum: VI	R Cerebellum: Crus 1	2.00	967	4.83	0.000*
16	−76	−32	R Cerebellum: Crus 1	R Cerebellum: Crus 2	2.83		4.16	0.001
8	−76	−32	R Cerebellum: Crus 1	Vermis	2.00		4.16	0.001
−14	−24	−2	L Thalamus	L Hippocampus	10.77	275	4.13	0.001
−4	−22	4	L Thalamus	R Thalamus	6.00		3.92	0.001
−14	−30	−12	L Parahippocampal gyrus	L Hippocampus	2.83	61	3.94	0.001
−42	−38	26	L Supramarginal gyrus	L STS	4.00	43	3.82	0.002
8	14	34	R Middle Cingulate gyrus	R Anterior Cingulate gyrus	6.00	33	3.81	0.002
−62	0	26	L Postcentral gyrus	L Precentral gyrus	2.00	97	3.77	0.002
−28	−64	−34	L Cerebellum: Crus 1	L Cerebellum: VI	2.83	48	3.76	0.002
20	6	−2	R Pallidum	R Putamen	2.00	160	3.61	0.003
Rest > Active task								
14	−60	16	R Calcarine fissure	R PreCuneus	3.46	9194	5.31	0.000*
−10	−66	16	L Calcarine fissure	L Cuneus	4.47		5.25	0.000*
−10	−58	14	L PreCuneus	L Calcarine fissure	2.83		5.23	0.000*
−6	24	−18	L Rectus	L Medial frontal gyrus: Orbitalis	4.00	8605	5.18	0.000*
2	38	−8	R Medial frontal gyrus: Orbitalis	R Anterior Cingulate gyrus	2.00		4.99	0.000*
6	54	−14	R Medial frontal gyrus: Orbitalis	R Rectus gyrus	2.00		4.90	0.000*
−30	−20	−26	L FG	L Parahippocampal gyrus	2.83	801	4.49	0.001
−24	−40	−28	L Cerebellum: IV/V	L Cerebellum: VI	4.47		4.24	0.001
−28	−42	−18	L FG	L Cerebellum: IV/V	4.47		4.20	0.001

Note: Cluster size is based on a voxel-wise threshold of Fdr-corrected $P < 0.01$. Local maxima more than 8.0 mm apart reported. *Denotes $P < 0.05$ after family-wise error correction. Automatic anatomical labeling of the peak voxel and the nearest region for the respective volume was based on Tzourio-Mazoyer et al. (2002).

Table 2

Rhyme focus > melodic focus activations

Anatomical location		MNI coordinates			Statistical values		
Region	BA	*x*	*y*	*z*	*N* voxels	Z	$P_{Fdr\text{-}corr}$
Left IFG	47	−40	38	−10	2503	5.54	0.002
	47	−42	26	−8		5.13	0.004
	38	−54	22	−12		4.71	0.005
Left FG	19	−42	−64	−16	191	3.91	0.016
Left inferior temporal gyrus	37	−56	−44	−16	33	3.89	0.016
Left MTG	21	−62	−24	−4	69	3.51	0.023
Right STS	22	66	−16	−6	14	3.42	0.027
Cerebellum		18	−76	−42	45	4.00	0.015

Note: Cluster size is based on a voxel-wise threshold of Fdr-corrected $P < 0.05$. Local maxima more than 8.0 mm apart reported. BA = Brodmann Area.

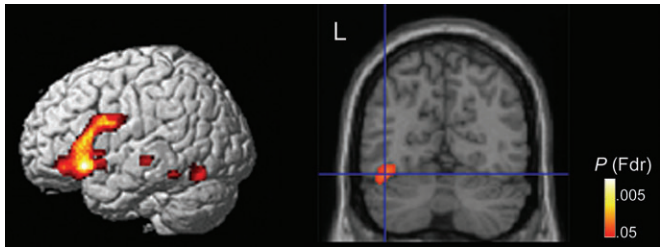

Figure 2. Rhyme focus > melodic focus condition activations. Selectively attending to speech, relative to selectively attending to melody, leads to increased activity in left inferior frontal regions, left mid-FG in the vicinity of the VWFA (coronal view on the right panel, $y = -63$), as well as clusters in temporal areas. Voxel threshold: Fdr-corrected $P < 0.05$. For a full list of activated regions and statistics, see Table 2.

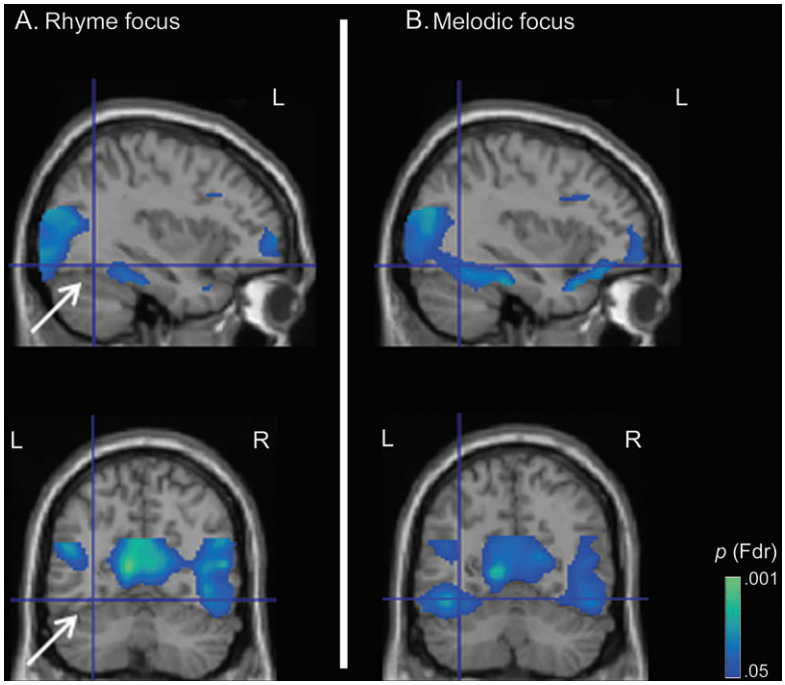

Figure 3. Illustration of extensive deactivations in extrastriate regions under rhyme focus (*A*) and under melodic focus (*B*). Notably, mid-FG (white arrow, panel *A*) is only deactivated when selectively attending to melody and not when selectively attending to speech. Deactivations along the entire anterior–posterior extent of the left FG are present only under melodic focus (top row). Further, mid-FG is equivalently deactivated in both left and right hemispheres only under melodic focus (bottom row). Rest > active (rhyme/melodic) condition. Voxel threshold: Fdr-corrected $P < 0.05$ (top: $x = -35$, bottom: $y = -58$).

Table 3

Statistical comparisons (paired *t*-tests) of rest > rhyme focus and rest > melodic focus for each fusiform ROI

	Left hemisphere			Right hemisphere		
	Anterior	Mid-	Posterior	Anterior	Mid-	Posterior
Rhyme focus	$t_{11} = 5.099$, $P < 0.0005$	$t_{11} = 1.597$, $P = 0.139$ ns	$t_{11} = 2.599$, $P < 0.05$	$t_{11} = 4.536$, $P < 0.001$	$t_{11} = 2.236$, $P < 0.05$	$t_{11} = 2.212$, $P < 0.05$
Melodic focus	$t_{11} = 6.467$, $P < 0.00005$	$t_{11} = 5.040$, $P < 0.0005$	$t_{11} = 2.840$, $P < 0.05$	$t_{11} = 4.112$, $P < 0.005$	$t_{11} = 2.913$, $P < 0.05$	$t_{11} = 2.270$, $P < 0.05$

each focus condition for every ROI. The resulting deactivation index (percent signal change) was used as a dependent measure in an omnibus 2 × 3 × 2 ANOVA with factors hemisphere (left, right), fusiform region (anterior, mid-, posterior), and top-down focus condition (rhyme, melodic). Note that this analysis provides a direct statistical test of whether deactivation is significantly greater in one ROI versus another based on top-down focus. The omnibus 2 × 3 × 2 ANOVA revealed a 3-way interaction between factors hemisphere, fusiform region, and focus condition ($F_{2,10} = 8.87$, $P < 0.01$). This interaction reflected the observation that melodic focus was associated with equivalent deactivation levels in all fusiform regions, whereas rhyme focus showed a differential pattern of top-down modulation across regions, characterized by a difference between left mid-fusiform and its neighboring fusiform regions in the left hemisphere (Fig. 4).

To investigate whether the 3-way interaction was driven by significant tuning effects of the left mid-fusiform ROI relative to other regions (topographically), we conducted a series of post hoc analyses testing specifically whether left fusiform deactivation in the rhyme focus condition was significantly different across the anterior, mid-, and posterior ROIs, and whether such topographic effects manifested in the melodic focus condition. Thus, we performed 1-way ANOVAs (using factor region with 3 levels: anterior, mid-, posterior fusiform ROI) separately for each hemisphere and for each attention focus condition. No evidence for a topographic effect was found during the melodic focus in the left hemisphere ($F_{2,10} = 0.801$, $P = 0.476$), or in right hemisphere for either condition (rhyme focus: $F_{2,10} = 2.039$, $P = 0.181$; melodic focus: $F_{2,10} = 0.649$, $P = 0.543$). The regional effect appeared only within the left hemisphere fusiform analysis during the rhyme focus condition ($F_{2,10} = 7.075$, $P < 0.05$). Further post hoc *t*-tests demonstrated that the left mid-fusiform ROI was significantly less deactivated compared to both the anterior ROI ($t_{11} = 3.427$, $P < 0.01$) and the posterior ROI ($t_{11} = 2.275$, $P < 0.05$). Overall, these findings suggest the rhyme focus modulation manifested as a form of topographic tuning, which was absent in the right hemisphere and the other attention condition.

Finally, given our *a priori* interest in the role of selective attention to phonology in basing mid-fusiform activity leftward (i.e., favoring the VWFA) we conducted a post hoc laterality analysis of left and right mid-fusiform activation. This took the form of a 2 × 2 ANOVA (focus condition: rhyme, melodic focus; hemisphere: left, right). Results indicated that differential deactivation based on linguistic focus demands manifested only in left mid-FG (hemisphere-by-task interaction $F_{1,11} = 7.003$, $P < 0.05$).

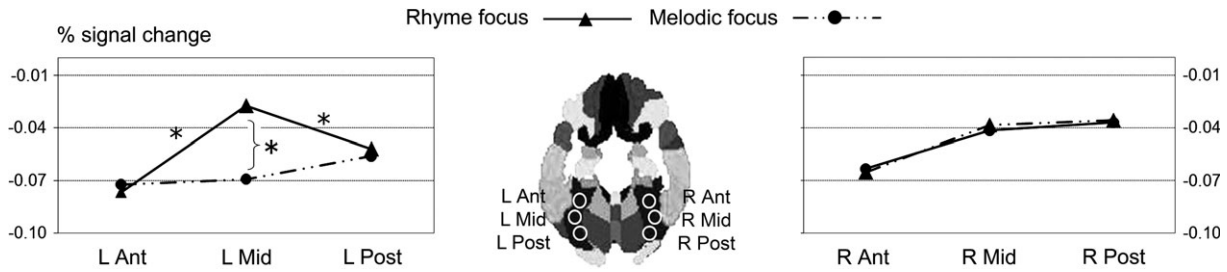

Figure 4. Deactivation patterns in ROIs in FG under rhyme and under melodic focus. Percent signal change for rhyme focus > rest and melodic focus > rest in the mid-FG/VWFA (−35, −58, −15; L Mid), surrounding anterior (−31, −42, −18; L Ant) and posterior (−30, −75, −14; L Post) fusiform ROIs, and homologues in the right hemisphere: (34, −41, −18; R Ant), (33, −58, −14; R Mid), and (30, −75, −12; R Post), respectively. Selective attention to speech modulates activity in fusiform regions in a topographic fashion, such that VWFA exhibits a peak activity relative to surrounding regions in the left hemisphere. Such attentional topographic effects are not present in the right hemisphere. Anatomical ROI locations (diagram showing locations in middle panel) were chosen based on Tzourio-Mazoyer et al. (2002).

Discussion

This study demonstrates that selective auditory attention to phonological versus melodic aspects of complex sounds drives patterns of differential blood oxygenation level–dependent (BOLD) activity in left mid-FG, left inferior frontal and bilateral temporal regions. Notably, the effect is observed under 2 conditions of identical bottom-up stimulation with active listening demands leading to equivalent behavioral performance. In light of the experimental design, which manipulated only processing goals such that attention was focused on linguistic versus melodic analysis, we interpret the present findings as reflecting the impact of top-down attention to language on cortical responses to speech sounds.

The interplay of top-down and bottom-up attentional processes can be considered with respect to the key brain regions involved: the prefrontal cortex, which represents goals and the means to achieve them, as it exerts top-down control (Posner and Petersen 1990; Miller and Cohen 2001) typically over perceptual areas, which exhibit response sensitivity to bottom-up stimulus properties. Recent investigations of language processing have successfully employed this construct. For instance, selectively attending to phonological versus orthographic aspects of written words has been shown to enhance the modulatory influence of IFG over task-specific areas, in line with the notion that prefrontal cortex sets the cognitive context relevant to particular processing goals through top-down projections to regions selective for carrying out the respective task demands (Bitan et al. 2005).

Numerous neuroimaging findings have associated different aspects of language processing with activations in particular cortical regions. Below we consider the profile of the attentional effect in the clusters that were differentially activated under rhyme versus melodic focus in the context of their functional involvement in processing linguistic content.

Frontal Areas

The left-lateralized linguistic focus effect in IFG is consistent with the routinely reported engagement of inferior frontal areas in language tasks (Demonet et al. 2005; Vigneau et al. 2006). The increased left IFG activation during rhyming relative to a control task (Paulesu et al. 1993; Booth et al. 2002) might be linked to speech stream segmentation into phonemes or syllables (Burton et al. 2000; Sanders et al. 2002). In addition to phonological processes, rhyming could also involve retrieval of semantic representations, as implied by the extent of the present linguistic IFG effect spanning across functionally heterogeneous ventral and dorsal IFG regions (Poldrack et al. 1999). In the framework of theories supporting motor system participation in speech recognition (Guenther and Perkell 2004; Skipper et al. 2005; Galantucci et al. 2006) engagement of left premotor and left opercular IFG areas during rhyming could reflect activation of motor representation for auditory words. The observed linguistic effect also fits with the proposed role of left premotor regions in subserving phonological short-term memory, which may be relevant during rhyming (Hickok and Poeppel 2007). Another account of the left IFG modulation, related to, but not specific to linguistic processing, is that rather than engaging perceptual representations per se, the 2 tasks differentially engage their associated action (or articulatory) codes. An individual's prior motor experience with the stimulus was not explicitly controlled for, leaving open the possibility that regions activated by previously formed action–sound representations (i.e., articulatory speech codes versus potentially absent action codes for the tones; Lahav et al. 2007) might have also contributed to the present left IFG task modulation.

Temporal Cortex

Across both active listening contexts, processing of the complex auditory stimuli elicited robust, extensive activations in lateral temporal cortices, in line with the central role of temporal cortex in sound analysis and speech perception (Zatorre et al. 2004; Demonet et al. 2005). Such activation patterns, independent of processing goals, were expected given the sensitivity of temporal regions to speech-like sounds in the absence of explicit focus on speech (Zevin and McCandliss 2005), and even awareness or consciousness (Davis et al. 2007). Responses in temporal cortex that were specific to the linguistic attentional focus, on the other hand, were restricted to 3 relatively small clusters located in left ITS, left MTG, and right STG/STS. Relevant to rhyming, mid-posterior STS areas might have been recruited as part of a network involved in phonological-level processing and representation, whereas left ITS might have been activated in its posited capacity of a lexical interface linking phonological and semantic information (Hickok and Poeppel 2007).

Evidence that selective auditory attention to language modulated specific regions in temporal cortex was not very robust. Two types of factors could have contributed to this end: the saturation of the BOLD response and the complex nature of the attentional effects. The acoustically challenging scanning environment along with the active listening demands could

have produced a ceiling effect in the BOLD measure, thus reducing fine-grained distinctions in the responses of auditory regions. Electrophysiological studies of the human auditory cortex have revealed that selective auditory attention to concurrent sounds operates through the interplay of facilitation of goal-relevant sound aspects and inhibition of irrelevant ones (Bidet-Caulet et al. 2007). The likely involvement of such opposing attentional influences (also given the current challenging perceptual demands; Lavie 2005) might have prevented better resolution of the linguistic effect in auditory regions. Overall, temporal areas exhibited robust activations in both active listening conditions, and modulations, albeit modest, by linguistic processing goals, in line with the notion that responses in language-sensitive perceptual areas are subject to tuning by attentional mechanisms.

Extrastriate Regions and FG

Interestingly responses in extrastriate regions during both active listening conditions were generally characterized by a decrease in BOLD signal relative to fixation baseline. Consideration of deactivations can be differentiated into: 1) those potentially related to a *default network* (Raichle et al. 2001), which should co-localize with a broad number of rest > task activation patterns reported in the literature; 2) those related to sensory suppression, which during a challenging auditory task should largely manifest in visual regions not typically involved in the default network; and 3) those that specifically differ across the 2 auditory attention tasks, which we have isolated to left mid-FG.

The functional significance of a BOLD decrease can also be regarded in light of this 3-fold categorization. Default network deactivations are typically proposed to reflect relative increases in complementary processes most active after completing a challenging active task (Gusnard et al. 2001). Indeed, the magnitude of deactivation in the default network has been shown to co-vary with the degree of task difficulty as assessed by auditory stimulus discriminability, stimulus rate presentation, and short-term memory load (McKiernan et al. 2003). In the context of these findings, our results of equivalent deactivation levels in default network regions across the 2 attention focus conditions is at least consistent with the behavioral metrics indicative of equated task difficulty. Task-dependent sensory deactivations that fall beyond the extent of the default network (such as our extrastriate BOLD decrease) have been more directly linked to the degree of processing from competing sensory modalities (Kawashima et al. 1995; Laurienti et al. 2002), and could reflect inhibition of the deactivated regions (Shmuel et al. 2006). In this sense, the overall deactivation of extrastriate visual cortex when attention is directed to auditory signals requiring demanding judgments and the lack of differential deactivation in the default network during equally challenging tasks are not surprising. What is remarkable in the present study is the emergence of a distinct, third type of phenomenon: the regionally specific pattern driven by attentional focus to phonological information that selectively spares left mid-FG from the extensive deactivation present under the equally difficult melodic attentional focus. This finding suggests that cross-modal attentional mechanisms may be sensitive to the linguistic nature of the processing goals.

The current fMRI results indicate that selective auditory attention to speech does not merely influence the degree of extrastriate deactivation, but rather impacts the topographic distribution of this deactivation, reflecting a form of top-down attentional topographic "tuning" of extrastriate activity in the service of processing different categories of information (i.e., phonological analysis of speech versus melodic analysis of tones). This topographic manifestation during selective attention to rhyming information is consistent with a distributed representational model of category selectivity within the ventral occipito-temporal cortex (Haxby et al. 2001) on a coarse, voxel-level scale (Haxby 2006). Top-down processes have been shown to modulate responses in distinct areas of extrastriate cortex pertinent to perception of particular visual stimulus features or visual categories (Chawla et al. 1999; O'Craven et al. 1999; Flowers et al. 2004). The present findings expand this notion to suggest an important role for top-down attention in driving topographic effects related to representations of different object categories (Haxby et al. 2001). The need for top-down attentional selection due to the competition between multiple stimulus dimensions (for discussion, see Desimone and Duncan 1995) and the lack of relevant visual information in our experiment have likely emphasized the tuning effect of top-down linguistic focus. Attentional factors might have been present, but not highlighted, in paradigms where participants were not explicitly focused on the relevant attribute for categorization, for example, passive fMRI adaptation techniques (Grill-Spector and Malach 2001).

But why does attention to speech content specifically produce a topographic tuning of left FG that favors recruitment of the mid-fusiform area, relative to anterior and posterior regions? The effect of selective auditory attention to speech falls in the vicinity of the VWFA, a region frequently engaged in reading (McCandliss et al. 2003a) and spelling tasks (Booth et al. 2002). Converging lines of evidence from studies with literate adults (Dehaene et al. 2004; Binder et al. 2006), lesion patients (Cohen et al. 2003), and developmental populations across fluency accruement (Shaywitz et al. 2002) have established that activity in the VWFA and neighboring regions functionally contributes to skilled reading. Notably, the present attentional effect in the left mid-FG occurred in the absence of visual stimulation, under identical auditory stimulation, and equated task difficulty. Thus, the topographic tuning of FG activity by attentional focus on speech could reflect activation of orthographic codes during demanding rhyming judgments. This interpretation is in line with the proposed involvement of this region in the integration of orthographic and phonological codes in proficient readers (Schlaggar and McCandliss 2007). Further research is necessary to demonstrate computational overlap between the current effects and those related to visual word form reading (for discussion, see Poldrack 2006). Additionally, given the temporal limitations of fMRI, it is difficult to assess how the reported selective attention effects relate to initial stimulus encoding, comparison, response execution, or post-comparison evaluation processes. To address this issue, we studied the same paradigm through the excellent temporal resolution of electroencephalography. Selective attention to speech showed an impact on event-related potentials during the perception of both the first and second words of the pair, indicating that top-down focus modulates early perceptual encoding (Yoncheva et al. 2008).

In a broader context, orthographic influences on spoken word perception have been reported across a gamut of linguistic processing goals (e.g., from phoneme and syllable monitoring to lexical decision; for discussion, see Ziegler et al.

2003). It is also plausible that—in the challenging acoustic context of the current paradigm—word recognition processes utilize all relevant information, thus recruiting associated orthographic representations. This takes place even though spelling is neither explicitly required nor necessarily beneficial for performing a rhyme judgment. Hence, the present effect, which is unlikely to be restricted to specific rhyming demands, potentially reflects a more general phenomenon when attending to linguistic content. Findings that selective auditory attention to speech sounds in dyslexic adults produces patterns of deactivation in occipital areas that differ significantly from these observed in normal readers (Dufor et al. 2007) are also in agreement with such a conceptualization.

In sum, the current investigation demonstrates how top-down attentional focus on language impacts fMRI-BOLD responses when processing spoken words. Selective auditory attention to speech content modulates activity in VWFA, potentially indicating the integration of phonological and orthographic processes in the absence of visual word stimulation. Furthermore, the linguistic attentional effect in extrastriate cortex manifests as topographically specific patterns of deactivation, which might constitute a mechanism for top-down systems to bias posterior perceptual networks.

Broader Implications

This approach to isolating the impact of top-down selective auditory attention to phonological information may prove valuable for future investigations into how attention to phonology influences reading acquisition and the rise of functional specialization of the VWFA. For instance, it is likely that individual differences in the ability to attend to phonological information associated with word spellings contribute to developmental reading disabilities (for review, see Schlaggar and McCandliss 2007). Recent developmental studies in fact have demonstrated that tasks involving phonological analysis of auditory words tend to activate VWFA increasingly across development and literacy skill acquisition (Booth et al. 2007; Cone et al. 2008). Future research isolating the role of selective attention to phonology may prove critical in demonstrating the importance of such attentional mechanisms in the development of functional specialization of the VWFA. As such, experimental training studies that manipulate the degree to which learners selectively attend to phonological and orthographic information reveal that this form of selective attention may be a key modulator of both functional reorganization of VWFA responses and success in reading acquisition (e.g., McCandliss et al. 2003b; Yoncheva et al. forthcoming). Thus, an understanding of the specific impact of selective attention to phonological information may prove critical to illuminating the neural mechanisms at play in the process of acquiring literacy.

Funding

National Institutes of Health (NIDCD R01 DC007694 PI) to B.D.M; National Science Foundation (REC-0337715 PI) to B.D.M.

Notes

We would like to thank Dr Jeremy Skipper for his valuable feedback and Dr Eva Hulse for her help with editing. *Conflict of Interest*: None declared.

Address correspondence to Bruce D. McCandliss, PhD, Box 552, 230 Appleton Place, Vanderbilt University, Nashville, TN 37203-5721, USA. Email: bruce.mccandliss@vanderbilt.edu.

References

Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O. 2007. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. J Neurosci. 27:9252–9261.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 10:512–528.

Binder JR, Medler DA, Westbury CF, Liebenthal E, Buchanan L. 2006. Tuning of the human left fusiform gyrus to sublexical orthographic structure. Neuroimage. 33:739–748.

Bitan T, Booth JR, Choy J, Burman DD, Gitelman DR, Mesulam MM. 2005. Shifts of effective connectivity within a language network during rhyming and spelling. J Neurosci. 25:5397–5403.

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.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM. 2003. Relation between brain activation and lexical performance. Hum Brain Mapp. 19:155–169.

Booth JR, Cho S, Burman DD, Bitan T. 2007. Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. Dev Sci. 10:441–451.

Bradley L, Bryant PE. 1983. Categorizing sounds and learning to read—a causal connection. Nature. 301:419–421.

Bregman AS. 1990. Auditory scene analysis: the perceptual organization of sound. Cambridge (MA): MIT Press.

Brem S, Halder P, Bucher K, Summers P, Martin E, Brandeis D. 2009. Tuning of the visual word processing system: distinct developmental ERP and fMRI effects. Hum Brain Mapp. 30:1833–1844.

Brett M, Anton J-L, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan Available on CD-ROM in NeuroImage, Vol 16, No 2.

Burton MW, Small SL, Blumstein SE. 2000. The role of segmentation in phonological processing: an fMRI investigation. J Cogn Neurosci. 12:679–690.

Chawla D, Rees G, Friston KJ. 1999. The physiological basis of attentional modulation in extrastriate visual areas. Nat Neurosci. 2:671–676.

Cohen L, Dehaene S. 2004. Specialization within the ventral stream: the case for the visual word form area. Neuroimage. 22:466–476.

Cohen L, Jobert A, Le Bihan D, Dehaene S. 2004. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage. 23:1256–1270.

Cohen L, Martinaud O, Lemer C, Lehericy S, Samson Y, Obadia M, Slachevsky A, Dehaene S. 2003. Visual word recognition in the left and right hemispheres: anatomical and functional correlates of peripheral alexias. Cereb Cortex. 13:1313–1333.

Cone NE, Burman DD, Bitan T, Bolger DJ, Booth JR. 2008. Developmental changes in brain regions involved in phonological and orthographic processing during spoken language processing. Neuroimage. 41:623–635.

Davis MH, Coleman MR, Absalom AR, Rodd JM, Johnsrude IS, Matta BF, Owen AM, Menon DK. 2007. Dissociating speech perception and comprehension at reduced levels of awareness. Proc Natl Acad Sci USA. 104:16032–16037.

Deese J. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. J Exp Psychol. 58:17–22.

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S. 2005. Neural correlates of switching from auditory to speech perception. Neuroimage. 24:21–33.

Dehaene S, Jobert A, Naccache L, Ciuciu P, Poline JB, Le Bihan D, Cohen L. 2004. Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. Psychol Sci. 15: 307–313.

Demonet JF, Price C, Wise R, Frackowiak RS. 1994. A PET study of cognitive strategies in normal subjects during language tasks. Influence of phonetic ambiguity and sequence processing on phoneme monitoring. Brain. 117:671–682.

Demonet JF, Thierry G, Cardebat D. 2005. Renewal of the neurophysiology of language: functional neuroimaging. Physiol Rev. 85:49–95.
Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. Annu Rev Neurosci. 18:193–222.
Dufor O, Serniclaes W, Sprenger-Charolles L, Demonet J-F. 2007. Top-down processes during auditory phoneme categorization in dyslexia: a PET study. Neuroimage. 34:1692–1707.
Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans AC, Zilles K, Amunts K. 2007. Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage. 36:511–521.
Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage. 25:1325–1335.
Faber A. 1992. Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In: Lima SD, Noonan M, Downing P, editors. The linguistics of literacy (typological studies in language). Amsterdam, Holland: John Benjamins. p. 111–134.
Flowers DL, Jones K, Noble K, VanMeter J, Zeffiro TA, Wood FB, Eden GF. 2004. Attention to single letters activates left extrastriate cortex. Neuroimage. 21:829–839.
Foxe JJ, Simpson GV. 2005. Biasing the brain's attentional set: II. Effects of selective intersensory attentional deployments on subsequent sensory processing. Exp Brain Res. 166:393–401.
Fritz JB, Elhilali M, David SV, Shamma SA. 2007. Auditory attention–focusing the searchlight on sound. Curr Opin Neurobiol. 17:437–455.
Galantucci B, Fowler CA, Turvey MT. 2006. The motor theory of speech perception reviewed. Psychon Bull Rev. 13:361–377.
Glover GH, Law CS. 2001. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med. 46:515–522.
Grainger J, Ferrand L. 1996. Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. J Mem Lang. 35:623–647.
Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst). 107:293–321.
Guenther FH, Perkell JS. 2004. A neural model of speech production and supporting experiments. Paper presented at 'From Sound to Sense' Conference. [cited 2008 Dec 15]. Available at: http://wwwrlemitedu/soundtosense/.
Gusnard DA, Raichle ME, Raichle ME. 2001. Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci. 2:685–694.
Haxby JV. 2006. Fine structure in representations of faces and objects. Nat Neurosci. 9:1084–1086.
Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.
Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL. 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci. 14:6336–6353.
Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.
Hillyard SA, Hink RF, Schwent VL, Picton TW. 1973. Electrical signs of selective attention in the human brain. Science. 182:177–180.
Kawashima R, O'Sullivan BT, Roland PE. 1995. Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the "mind's eye". Proc Natl Acad Sci USA. 92:5969–5972.
Lahav A, Saltzman E, Schlaug G. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci. 27:308–314.
Laurienti PJ, Burdette JH, Wallace MT, Yen YF, Field AS, Stein BE. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J Cogn Neurosci. 14:420–429.
Lavie N. 2005. Distracted and confused?: selective attention under load. Trends Cogn Sci. 9:75–82.
Luck SJ, Hillyard SA. 1995. The role of attention in feature detection and conjunction discrimination: an electrophysiological analysis. Int J Neurosci. 80:281–297.
McCandliss BD, Beck IL, Sandak R, Perfetti CA. 2003b. Focusing attention on decoding for children with poor reading skills: design and preliminary tests of the word building intervention. Sci Stud Reading. 7:75–104.
McCandliss BD, Cohen L, Dehaene S. 2003a. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn Sci. 7: 293–299.
McClelland JL, Rumelhart DE. 1981. An interactive activation model of context effects in letter perception: Part I. An account of basic findings. Psychol Rev. 88:375–407.
McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. 2003. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. J Cogn Neurosci. 15:394–408.
Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 24:167–202.
Morais J, Bertelson P, Cary L, Alegria J. 1986. Literacy training and speech segmentation. Cognition. 24:45–64.
O'Craven KM, Downing PE, Kanwisher N. 1999. fMRI evidence for objects as the units of attentional selection. Nature. 401:584–587.
Paulesu E, Frith CD, Frackowiak RS. 1993. The neural correlates of the verbal component of working memory. Nature. 362:342–345.
Poldrack RA. 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci. 10:59–63.
Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage. 10:15–35.
Port R. 2007. How are words stored in memory? Beyond phones and phonemes. New Ideas Psychol. 25:143–170.
Posner MI, Petersen SE. 1990. The attention system of the human brain. Annu Rev Neurosci. 13:25–42.
Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. 2001. A default mode of brain function. Proc Natl Acad Sci USA. 98:676–682.
Sanders LD, Newport EL, Neville HJ. 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat Neurosci. 5:700–703.
Schlaggar BL, McCandliss BD. 2007. Development of neural systems for reading. Annu Rev Neurosci. 30:475–503.
Seidenberg MS, McClelland JL. 1989. A distributed, developmental model of word recognition and naming. Psychol Rev. 96:523–568.
Seidenberg MS, Tanenhaus MK. 1979. Orthographic effects in rhyme monitoring. J Exp Psychol [Hum Learn]. 5:546–554.
Shaywitz BA, Shaywitz SE, Pugh KR, Mencl WE, Fulbright RK, Skudlarski P, Constable RT, Marchione KE, Fletcher JM, Lyon GR, et al. 2002. Disruption of posterior brain systems for reading in children with developmental dyslexia. Biol Psychiatry. 52:101–110.
Shmuel A, Augath M, Oeltermann A, Logothetis NK. 2006. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nat Neurosci. 9:569–577.
Skipper JI, Nusbaum HC, Small SL. 2005. Listening to talking faces: motor cortical activation during speech perception. Neuroimage. 25:76–89.
Sommers MS, Lewis BP. 1999. Who really lives next door: Creating false memories with phonological neighbors. J Mem Lang. 40:83–108.
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.
Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage. 30:1414–1432.
Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L. 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron. 55: 143–156.

Woodcock RW, McGrew K, Mather N. 2001. Woodcock-Johnson III. Itasca (IL): Riverside Publishing.
Xue G, Poldrack RA. 2007. The neural substrates of visual perceptual learning of words: Implications for the visual word form area hypothesis. J Cogn Neurosci. 19:1643–1655.
Yoncheva YN, Blau VC, Maurer U, McCandliss BD. Forthcoming. N170 in learning to read a novel script: the impact of attending to phonology on lateralization. Dev Neuropsychol.
Yoncheva YN, Zevin JD, Maurer U, McCandliss BD. 2008. The temporal dynamics of listening to versus hearing words: attention modulates both early stimulus encoding and preparatory activity. J Cogn Neurosci. Supplement ISSN 1096-8857:268.
Zatorre RJ, Bouffard M, Belin P. 2004. Sensitivity to auditory object features in human temporal neocortex. J Neurosci. 24: 3637–3642.
Zevin JD, McCandliss BD. 2005. Dishabituation of the BOLD response to speech sounds. Behav Brain Funct. 1:4.
Ziegler JC, Muneaux M, Grainger J. 2003. Neighborhood effects in auditory word recognition: phonological competition and orthographic facilitation. J Mem Lang. 48:779–793.