Behavioral/Cognitive

Selective and Invariant Neural Responses to Spoken and Written Narratives

Mor Regev,[1,2] Christopher J. Honey,[1,2] Erez Simony,[1,2] and Uri Hasson[1,2]
[1]Department of Psychology and [2]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08540

Linguistic content can be conveyed both in speech and in writing. But how similar is the neural processing when the same real-life information is presented in spoken and written form? Using functional magnetic resonance imaging, we recorded neural responses from human subjects who either listened to a 7 min spoken narrative or read a time-locked presentation of its transcript. Next, within each brain area, we directly compared the response time courses elicited by the written and spoken narrative. Early visual areas responded selectively to the written version, and early auditory areas to the spoken version of the narrative. In addition, many higher-order parietal and frontal areas demonstrated strong selectivity, responding far more reliably to either the spoken or written form of the narrative. By contrast, the response time courses along the superior temporal gyrus and inferior frontal gyrus were remarkably similar for spoken and written narratives, indicating strong modality-invariance of linguistic processing in these circuits. These results suggest that our ability to extract the same information from spoken and written forms arises from a mixture of selective neural processes in early (perceptual) and high-order (control) areas, and modality-invariant responses in linguistic and extra-linguistic areas.

Introduction

Until ~5000 years ago, before the development of logographic and alphabetic writing systems, human language relied mainly upon spoken utterances (Houston, 2004). The advent of written language provided a new, visual pathway for communication. However, because written language requires extensive training and typically follows the acquisition of spoken language, it is thought to rely on neural pathways that originally supported spoken language (van Atteveldt et al., 2004). But which brain regions are common to the written and spoken language systems, and do they function in the same way across modalities?

Prior work has mapped the regions responsive to linguistic stimuli presented visually (writing) and auditorily (speech). Regions that show "modality-selective" responses were defined as those that produce aggregate activity increases for stimuli of only a single modality. Regions that show "modality-invariant" responses were defined as those that responded to both written and spoken stimuli. This usage of "modality" therefore subsumes both the sensory modality (auditory vs visual) and the task modality (listening vs reading). Modality selectivity was observed in sensory cortices: early auditory cortex responds to spoken (but not written) stimuli, whereas early visual cortex responds to written (but not spoken) stimuli. Modality-invariant activations were observed in widespread language systems, including the posterior superior temporal cortex, inferior parietal cortex, and subsets of inferior frontal cortex (Raij et al., 2000; Shaywitz et al., 2001; Booth et al., 2002; Marinković, 2004; van Atteveldt et al., 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012). Thus, prior studies suggest a hierarchical model in which early sensory regions are modality-selective but the written and spoken systems gradually converge, so that modality-invariance increases toward higher-order language regions.

The evidence supporting the hierarchical convergence of spoken and written systems can be criticized in two respects. First, responses to language stimuli were measured via aggregate activations to constrained stimuli such as isolated words or isolated short sentences. These paradigms do not map the full set of regions engaged in real-life language comprehension (Lerner et al., 2011; Ben-Yakov et al., 2012), and thus may underestimate the responses in high-order regions to spoken and written language, as well as their overlap. Second, the demonstration of spatially overlapping neural responses is not strong evidence for modality invariance, because an aggregate activation may be observed in both modalities even when different kinds of processing are taking place (Dinstein et al., 2007; Ben-Yakov et al., 2012; Honey et al., 2012). Although spatial overlap between averaged signals is an informative finding that suggests some level of modality invariance, a stronger form of modality invariance is indicated when a region responds with the same temporal response profile to spoken and written forms of the same linguistic input.

In the current study, we measured temporal response profiles to 7 min real-life narrative stimuli and directly compared the response time courses within and across the spoken and written versions of the narrative. Real-life linguistic stimuli can evoke highly reliable and selective responses, even in regions that show little response modulation to isolated letters, words, or sentences

Received April 13, 2013; revised Aug. 13, 2013; accepted Aug. 15, 2013.

Author contributions: M.R. and U.H. designed research; M.R. performed research; C.J.H. and E.S. contributed unpublished reagents/analytic tools; M.R. analyzed data; M.R., C.J.H., and U.H. wrote the paper.

This work was supported by NIH Grant R01-MH094480 (to U.H., M.R., C.J.H., and E.S.). We thank Ido Davidesco and Janice Chen for helpful comments on the paper, and Yulia Lerner for sharing her data.

The authors declare no competing financial interests.

Correspondence should be addressed to Uri Hasson, 3-C-13 Green Hall, Psychology Department, Princeton University, Princeton, NJ 08540. E-mail: hasson@princeton.edu.

DOI:10.1523/JNEUROSCI.1580-13.2013

 0270-6474/13/3315978-11$15.00/0

(Lerner et al., 2011). Intersubject correlation analysis (inter-SC; Hasson et al., 2010) provides an ideal tool for measuring the reliability of response time course to natural stimuli. Each subject in this study either heard or viewed a spoken or a written version of the same continuous narrative while undergoing functional magnetic resonance imaging (fMRI). Subjects were instructed to attend to the details of the narrative, and a postscan questionnaire assessed their comprehension and engagement.

This design allowed us to identify two types of response profiles: (1) modality-selective responses in areas which responded more reliably across subjects to the written (or spoken) narrative, and (2) potentially modality-invariant responses in areas which responded equally reliably to the spoken and written narratives. For each potentially modality-invariant region, we then examined whether that region produced the same time-varying response profile when the narrative was spoken and when it was written by performing intersubject correlations across modalities. Regions passing this test were classified as truly displaying modality-invariant response. Throughout the paper, the term modality refers to both the "sensory modality" (auditory vs visual) as well as the "task modality" (listening vs reading). Using this approach, we identified robust modality-invariant responses in linguistic areas along the posterior superior temporal gyrus (pSTG) and in the left inferior frontal gyrus, as well as in some extra-linguistic areas, such as the precuneus. However, not all higher-order areas demonstrated modality-invariant responses: some parietal and frontal areas produced responses that were selective for either the spoken or written stories, in addition to the selectivity observed in early sensory areas.

Materials and Methods

Subjects

Thirty-eight subjects successfully participated in one of the two main experimental conditions (written narrative and spoken narrative), or in a third condition (combined narrative), designed to guide us in defining a set of independent regions of interest (ROIs). Eleven subjects were discarded from the analysis: four subjects due to head motion >2 mm, two due to corrupted functional signal, one due to corrupted anatomical signal, one due to anomalous anatomy, one due to difficulties in hearing the stimulus, and two due to failure of the stimulus comprehension test. Additional subjects were scanned until data from nine subjects were collected for spoken (four males, five females; ages 19–28), written (five males, four females; ages 19–22), and combined (five males, four females; ages 19–22) narrative conditions. In addition, nine of the subjects from the written and spoken conditions also participated in an unintelligible written control experiment, and an additional set of 11 subjects participated in an unintelligible spoken control experiment. Procedures were approved by the Princeton University Committee on Activities Involving Human Subjects. All subjects were right-handed native-English speakers with normal hearing and provided written informed consent.

MRI acquisition

Subjects were scanned in a 3T full-body MRI scanner (Skyra, Siemens) with a 12-channel head coil. For functional scans, images were acquired using a T2*-weighted echo planer imaging (EPI) pulse sequence [repetition time (TR), 1500 ms; echo time (TE), 28 ms; flip angle, 64°], each volume comprising 27 slices of 4 mm thickness with 0 mm gap; slice acquisition order was interleaved. In-plane resolution was $3 \times 3\ mm^2$ [field of view (FOV), $192 \times 192\ mm^2$]. Anatomical images were acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) pulse sequence (TR, 2300 ms; TE, 3.08 ms; flip angle 9°; $0.89\ mm^3$ resolution; FOV, $256\ mm^2$). To minimize head movement, subjects' heads were stabilized with foam padding. Stimuli were presented using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). Subjects were provided with an MRI compatible in-ear mono earbuds (Sensimetrics model S14), which provided the same audio input to each ear. MRI-safe passive noise-canceling headphones were placed over the earbuds for noise removal and safety.

Stimuli and experimental design

The spoken language stimulus was a 7 min real-life story ("Pie-man," told by Jim O'Grady) recorded at a live storytelling performance ("The Moth" storytelling event, New York City). The written language stimulus was a 954-word transcript of the same narrative. The spoken and written versions of the narrative were combined simultaneously to create an audio-visual stimulus. The combined audiovisual experiment was used to define an unbiased set of ROIs. These three stimuli (Fig. 1) were presented in a between-subjects design; each subject participated in only one of the following conditions: the spoken condition (auditory stimulus), the written condition (visual stimulus), or the combined condition (audiovisual stimuli).

In the written condition, words were individually presented in the center of the screen in rapid serial visual presentation in a rhythm that accurately matched the timing of the original spoken version. In cases where a few spoken words were inseparable in time (46.17% of the screens), we presented a few words on the screen (two-word screens appeared 207 times, three-word screens 64 times, and four-word or more screens six times). Overall, the narrative contained 600 screen images, 0.7 ± 0.5 s each. Infrequently, the recording contained the laughter and applause of the audience. Each of these laughter/applause segments was classified as a "single word" event (5.5% of screens). A "smiley" icon was used in correspondence to these segments in the written condition. Neutral lead-in music was played for 12 s before the onset of the spoken stimulus, and graphical music symbols were shown for 12 s before the onset of the written stimulus. Responses to these initial 12 s were excluded from all analyses.

Subjects also participated in two control conditions, one for the spoken and one for the written conditions (Fig. 1). In the unintelligible spoken condition, the narrative waveform was played reversed in time (backward narrative), creating the perceptual effect of an unintelligible speech-like stimulus. In the unintelligible written condition, the letters constituting each word were randomly permuted and then the entire scrambled word was rotated 180 degrees, creating an unreadable array of unfamiliar letters from the exact same set of visual features.

Data analysis

Preprocessing. fMRI data were reconstructed and analyzed with the BrainVoyager QX software package (Brain Innovation) and with in-house software written in MATLAB (MathWorks). Preprocessing of functional scans included intrasession 3D motion correction, slice scan time correction, linear trend removal, and high-pass filtering (two cycles per condition). Spatial smoothing was applied using a Gaussian filter of 6 mm full-width at half-maximum value. The cortical surface was reconstructed from the 3D MPRAGE anatomical images using BrainVoyager software. The complete functional dataset was transformed to a 3D Talairach space (Talairach and Tournoux, 1988) and projected on a reconstruction of the cortical surface.

Inter-SC maps were produced for each condition (e.g., spoken narrative, written narrative, combined narrative) and across conditions (e.g., spoken narrative vs written narrative). The inter-SC maps provide a measure of the reliability of brain responses to each of the conditions by quantifying the correlation of the time course of BOLD activity across subjects listening to the spoken narrative or reading the same written narrative (Hasson et al., 2004, 2010; Lerner et al., 2011).

For each voxel, inter-SC within a condition is calculated as an average correlation $R = \frac{1}{N}\sum_{j=1}^{N} r_j$, where the individual r_j are the Pearson correlations between that voxel's BOLD time course in one individual and the average of that voxel's BOLD time courses in the remaining individuals. Inter-SC across conditions is calculated as an average $\tilde{R} = \frac{1}{N}\sum_{j=1}^{N} \tilde{r}_j$ over the correlations, $\tilde{r}_j$ between the BOLD time courses of the j'th individual from the first group and the average BOLD time courses of all the individuals in the other group. In a standard GLM analysis, experimenters usually assume prototypical response profile for

each specific stimulus. The inter-SC analysis method differs from conventional fMRI data analysis methods in that it circumvents the need to specify a model for the neuronal processes for any given condition. Instead, the inter-SC method uses the subject's brain responses within a given brain area (e.g., in the temporal parietal junction) as a model to predict brain responses to the same content. The inter-SC was calculated (1) within each condition (i.e., within the reading group and within the listening group) and (2) across conditions (i.e., across the reading and listening groups).

Projection of white matter. To diminish the impact of global, non-neural signal artifact on local BOLD signals, we projected-out the mean white matter signal from the BOLD signal in each voxel in each subject. The mean signal was calculated individually for each subject, and was entered into a linear regression to predict the BOLD signal in each voxel. The BOLD signals were then replaced with the residuals resulting from this regression, and the mean and variance of each of these residuals were matched to the mean and variance of the pre-projection BOLD signal.

Bootstrapping by phase-randomization. Because of the presence of long-range temporal autocorrelation in the BOLD signal (Zarahn et al., 1997), the statistical likelihood of each observed correlation was assessed using a bootstrapping procedure based on phase-randomization. The null hypothesis was that the BOLD signal in each voxel in each individual was independent of the BOLD signal values in the corresponding voxel in any other individual at any point in time (i.e., that there was no inter-SC between any pair of subjects).

For all conditions, a phase randomization of each voxel time course was performed by applying a fast Fourier transform to the signal, randomizing the phase of each Fourier component, and inverting the Fourier transformation. This procedure scrambles the phase of the BOLD time course but leaves its power spectrum intact. For each randomly phase-scrambled surrogate dataset, we computed the inter-SC (R) for all voxels in the exact same manner as the empirical correlation maps described above, i.e., by calculating the Pearson correlation between that voxel's BOLD time course in one individual and the average of that voxel's BOLD time courses in the remaining individuals. The resulting correlation values were averaged within each voxel across all subjects, creating a null distribution of average correlation values for all voxels.

To correct for multiple-comparisons, we selected the highest inter-SC value from the null distribution of all voxels in a given iteration. We repeated this bootstrap procedure 1000 times to obtain a null distribution of the maximum noise correlation values (i.e., the chance level of receiving high correlation values across all voxels in each iteration). Familywise error rate (FWER) was defined as the top 5% of the null distribution of the maximum correlations values exceeding a given threshold (R^*), which was used to threshold the veridical map (Nichols and Holmes, 2002). In other words, in the inter-SC map, only voxels with mean correlation value (R) above the threshold derived from the bootstrapping procedure (R^*) were considered significant after correction for multiple-comparisons and were presented on the final map. Using this method, the thresholds for each condition were as follows: spoken condition $R^* = 0.17$; written condition $R^* = 0.16$; combined condition $R^* = 0.15$; unintelligible speech control $R^* = 0.12$; unintelligible written control $R^* = 0.16$. The same procedure was performed on the correlation values which were computed between the spoken and the written conditions, producing a threshold of $\tilde{R}^* = 0.17$.

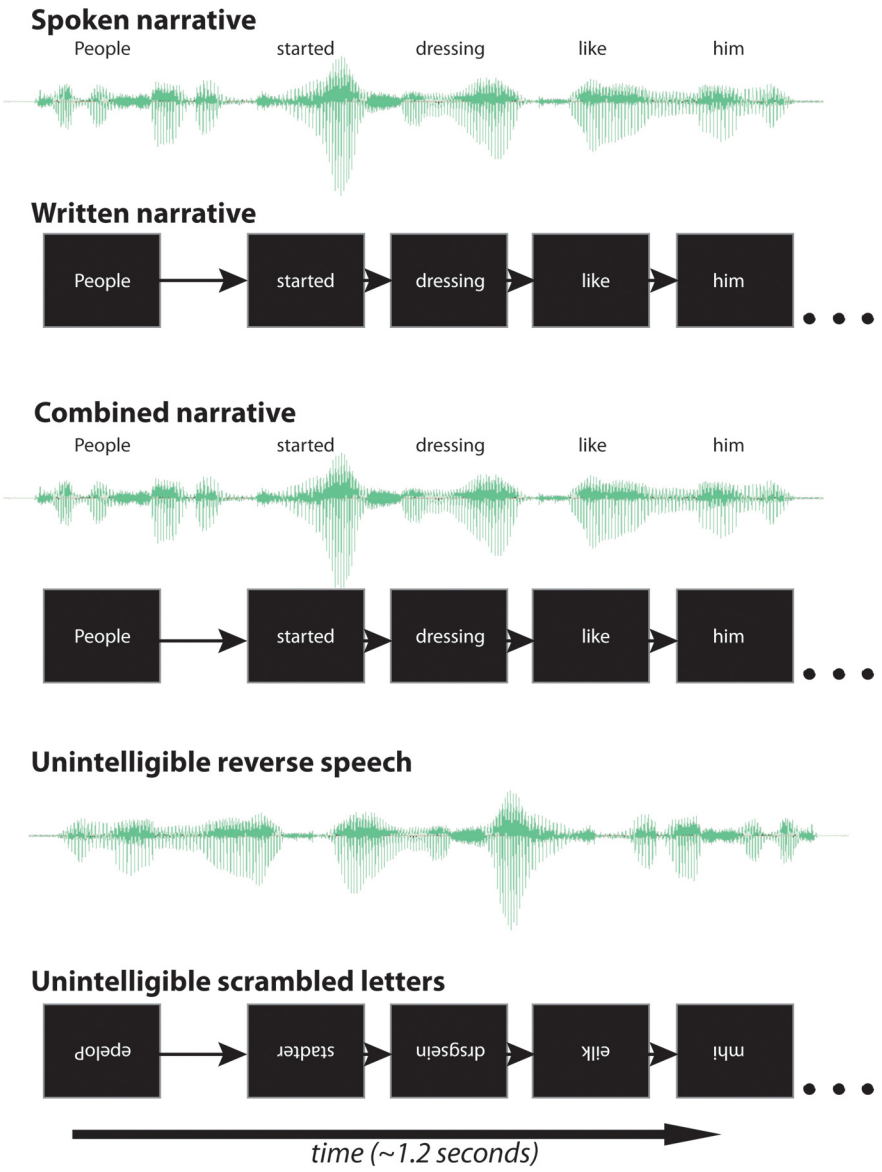

Figure 1. A 1.2 s segment of the 7 min narrative stimulus. While undergoing fMRI, subjects were exposed to one of two narrative presentation modes: a spoken version or a written version. Subjects were also exposed to a spoken and written control conditions. In the spoken control, the narrative was played backwards, creating a perceptual effect of unintelligible speech; in the written control, the letters in each word were permuted and the emerging letter array was rotated by 180 degrees, resulting in an unintelligible written stimulus.

To identify areas that show increase in response reliability for one condition over the other, a t test ($\alpha = 0.05$) was performed within each voxel that exceeded the threshold in at least one of the inspected conditions (see Fig. 4*A*). Thus, the t test was performed by comparing the correlation values of subjects from the spoken condition $\{r_j, r_{j+1} \ldots r_n\}$ to the correlation values of subjects from the written condition $\{r_j, r_{j+1} \ldots r_n\}$, within each voxel.

ROI analysis. In this work, we used two types of ROIs. (1) To sample an ROI in the primary auditory area (A1+), the narrative's sound envelope was convolved with a hemodynamic response function (Glover, 1999) to simulate a BOLD time course, and was used as a regressor. A bilateral superior temporal ROI called A1+ was then defined using two $10 \times 10 \times 10$ mm^3 cubes, located on the peaks of the audio regression in each hemisphere. (2) A set of independent ROIs (see Figs. 4, 5) was defined

Table 1. Talairach coordinates of independently defined ROIs

		Mean		
	Area	*x*	*y*	*z*
Left hemisphere	A1+	−47	−25	11
	pDLPFC	−42	4	29
	aDLPFC	−41	13	39
	aIPL	−49	−48	43
	IFG	−48	12	15
	pSTG	−54	−49	15
	Angular gyrus	−47	−59	18
Right hemisphere	A1+	48	−18	6.5
	aIPL	51	−43	42
	Angular gyrus	44	−64	30
Medial	V1+	8.7	−77	1.3
	Left precuneus	−2.6	−68	32
	Right precuneus	2.7	−61	32
	Left pdmPFC	−4.7	40	38
	Right pdmPFC	5.1	48	31

based on an intersubject reliability map calculated within an independent group of subjects who concurrently read and listened to the narrative (the combined condition). The ROIs were defined by sampling 252–3186 adjoining voxels around the response reliability peaks in the vicinity of the following areas: calcarine sulcus (primary visual area, V1+), left posterior dorsolateral prefrontal cortex (pDLPFC) and anterior dorsolateral prefrontal cortex (aDLPFC), the left inferior frontal gyrus (IFG) [which includes pars opercularis (approximately BA44) and pars triangularis (approximately BA45)], the left and right angular gyrus, left and right posterior regions of the dorsal medial prefrontal cortex (dmPFC), the left pSTG, the precuneus, and anterior regions in the left and right inferior parietal lobule (aIPL; Table 1).

To identify which of the ROIs show increase in response reliability for one condition over the other, a one tailed *t* test ($\alpha = 0.05$) was performed within each ROI. Thus, the *t* test was performed comparing the mean correlation values of all voxels within an ROI from all subjects at the spoken condition $\{\bar{r}_j, \bar{r}_{j+1} \ldots \bar{r}_n\}$ to the mean correlations value of all voxels within the ROI from all subjects at the written condition $\{\bar{r}_j, \bar{r}_{j+1} \ldots \bar{r}_n\}$.

Behavioral assessment

Immediately following the scan, we assessed each subject's engagement and the intelligibility of the stimulus using a simple questionnaire. Subjects were asked to write down a summary of the narrative they had just heard or read, as detailed as possible (spoken $n = 9$, written $n = 9$). Four independent raters graded these written records against a standard consisting of four questions about characters in the narrative, nine questions about particular events in the narrative, two questions about prominent keywords, as well as comprehensiveness level of the summary. The mean of the resulting scores (on a scale from 0 to 13) provided a measure of each subject's comprehension of the narrative. In addition, most of the subjects were asked to rate on a scale from 1 to 7 how difficult it was for them to reconstruct the narrative (spoken $n = 7$, written $n = 9$) and how engaged they felt with the narrative (spoken $n = 8$, written $n = 9$).

Two-tailed Welch's *t* tests ($\alpha = 0.05$) were conducted between-subjects to compare the effect of the different experimental conditions on self-reported engagement and recall difficulty, as well as independently rated narrative-comprehension.

Results

We compared the behavioral and neural responses within and between two groups of subjects. One group ("spoken") listened to a 7 min real-life spoken narrative, whereas the other group ("written") read an exact transcript of the spoken narrative, in which words were presented at the center of the screen, at a presentation rate which was matched to the spoken condition (see Materials and Methods; Fig. 1).

Behavioral results

Subjects comprehended the narrative well (spoken: M = 10, SD = 2.94; written: M = 8.86, SD = 2.32), with no difference in comprehension across the two conditions (spoken and written; $t_{(15.17)} = 0.86$, $p = 0.4$; Fig. 2*A*). The subjective level of difficulty recalling the narrative was low and equal across the two conditions ($t_{(13.99)} = -1$, $p = 0.33$; Fig. 2*B*). The subjective level of engagement was high and also equal across conditions ($t_{(8.57)} = -1.41$, $p = 0.19$; Fig. 2*C*). These results suggest that the nonstandard reading condition for the written group (see Materials and Methods) did not hinder their comprehension or engagement with the narrative, relative to the spoken group. We next compared the time courses of neural activation for spoken and written naturalistic language.

Identifying the language network involved in listening and reading

We began by identifying the set of brain areas that (1) responded reliably across subjects who listened to the spoken narrative or (2) responded reliably across subjects that read the written narrative. This was done by mapping the inter-SC separately for the subjects in the spoken group and for the subjects in the written group.

Intersubject correlation in the spoken condition

Consistent with previous reports (Lerner et al., 2011; Honey et al., 2012), the spoken condition showed reliable responses across subjects in early auditory areas, as well as linguistic and extralinguistic areas (Fig. 3*A*). Early auditory areas included primary and secondary cortices that process low-level features of the sound (A1+; Romanski and Averbeck, 2009). Linguistic areas include the pSTG and posterior superior temporal sulcus (pSTS), angular gyrus, supramarginal gyrus, posterior inferior parietal lobule, and IFG (which includes its opercular and ventral triangular parts; Table 2). Each of these regions has been previously linked with one or more core linguistic processes at the level of phonemes, lexical items, grammar, or articulation (Hickok and Poeppel, 2007; Sahin et al., 2009; Price, 2010). Finally, extralinguistic regions, which seem to be involved in processing the narrative and the social content of the story (Fletcher et al., 1995; Xu et al., 2005; Ferstl et al., 2008; Lerner et al., 2011), include the precuneus, the posterior cingulate cortex (PCC), left aDLPFC, left orbitofrontal cortex, and dmPFC (Table 2; summary of Talairach coordinates of all areas presented in each map).

Intersubject correlation in the written condition

Computing the inter-SC within the written condition (Fig. 3*B*) revealed reliable responses across subjects in the occipital visual cortex, as well as in many of the linguistic and extralinguistic areas observed in the spoken condition (Fig. 3*C*). Areas that exhibited a reliable response include the pSTG and pSTS, anterior superior temporal gyrus (aSTG), angular gyrus, and IFG, all in the language network, and extralinguistic areas including the precuneus, the aIPL, the PCC, the dmPFC, DLPFC, and the orbitofrontal cortex (Table 2).

Modality-selective responses

Next, considering all regions that responded reliably to one or both conditions, we sought to identify those voxels with significantly greater response reliability in one condition rather than another by using a voxelwise *t* test (see Materials and Methods).

The written condition evoked significantly greater reliability not only in the visual cortex, but also in the aIPL and some frontal areas including the left and right pDLPFC, dorsal regions in the

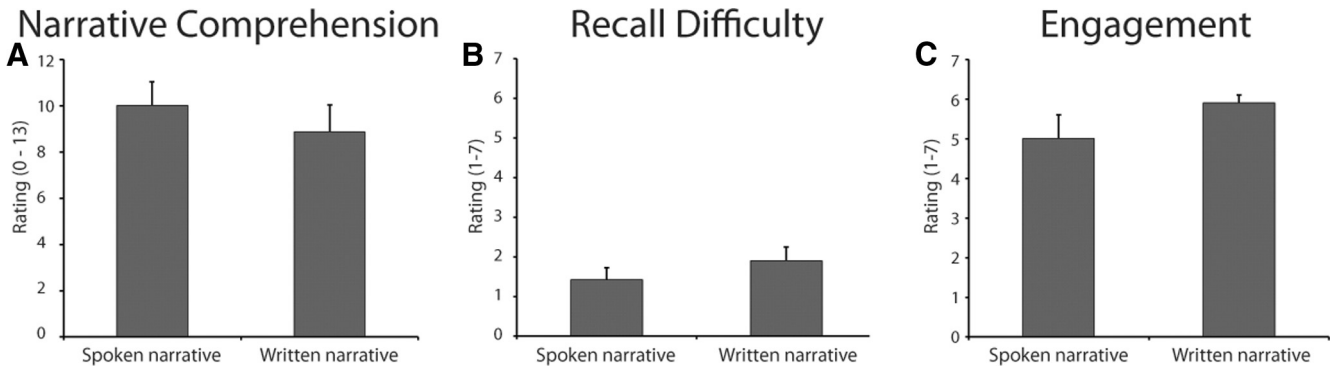

Figure 2. Behavioral measures of comprehension and engagement did not differ between the spoken and written conditions. ***A***, Comparable levels of narrative comprehension were measured across the two modes of narrative presentation. ***B***, Difficulty in recalling the narrative was equal across all conditions. ***C***, Similar levels of engagement were measured across the two modes of narrative presentation. Values are means and error bars represent SEM across subjects.

triangularis, the right orbitofrontal cortex, and posterior region within the right dmPFC (Fig. 4*A*, green; Table 2). The spoken condition evoked significantly greater reliability not only in early auditory cortex, but also in a smaller set of frontal and parietal areas, including the left anterior DLPFC, a posterior region within the left dmPFC, and right superior parietal lobule (Fig. 4*A*, red; Table 2). In addition, bilateral areas in the middle STG and an area within the right STS exhibited significantly more reliable response for the spoken condition than the written condition ($p < 0.05$). The responses in the right STS may be related to prosodic information, which is thought to be preferentially processed in the right hemisphere (Ross and Mesulam, 1979; George et al., 1996) and especially in the right STS (Belin et al., 2002; Bestelmeyer et al., 2011).

The results of the voxelwise analysis were reproduced in a group of sensory and high-order ROIs (see Materials and Methods). Although early auditory and visual modality-selective areas responded reliably to both intelligible speech and unintelligible scrambled stimuli, high-order parietal and frontal regions that show modality-selective responses did not respond reliably to the unintelligible scrambled stimuli (Fig. 4*B*). Early auditory areas (A1+), as defined using the narrative's acoustic envelope, showed reliable responses in the spoken condition, but not in the written condition ($t_{(16)} = 8.08$, $p \ll 0.0001$). Moreover, responses in A1+ were reliable even when the speech was unintelligible (played backwards), suggesting that this region is involved in low-level, prelinguistic processing of the spoken input. Unimodal visual areas (V1+), exhibited reliable responses in the written condition, but not in the spoken condition ($t_{(16)} = 11.65$, $p \ll 0.0001$). Moreover, this early visual area responded reliably, but to a lesser extent, when the letters in each word were scrambled and rotated to create unreadable arrays of visual input. This effect suggests that V1+ is involved in low-level processing of visual inputs, but may be influenced to some extent (via top-down feedback or attentional modulations) by the presence of readable orthographic input.

Some frontal ROIs such as the left posterior dmPFC exhibited significantly greater reliable response for the spoken condition relative to the written condition ($t_{(16)} = 3.03$, $p = 0.003$), but did not respond to the unintelligible written condition. Conversely, some high-order areas in the left pDLPFC ($t_{(16)} = 3.06$, $p = 0.004$) and the left and right aIPL($t_{(16)} = 5.02$, $p \ll 0.0001$; $t_{(16)} = 3.91$, $p = 0.001$) exhibited significantly greater reliable response for the written condition relative to the spoken condition, but did not respond to the unintelligible spoken condition. Overall, these results suggest differential involvement of these frontal areas in the processing of spoken and written information.

Modality-invariant responses

Spatial overlap of regions responsive to spoken and written narratives

Next we looked at areas that responded reliably to both the spoken and the written narratives. An overlap between the high-order cortical areas which responded reliably to both conditions was seen in many linguistic and extra-linguistic areas (Fig. 3*C*). The linguistic areas include the pSTG and pSTS, the angular gyrus, the supramarginal gyrus, and the IFG (which includes its opercular and ventral triangular parts). The extra-linguistic areas include the precuneus, PCC, anterior regions within the dmPFC, and the posterior IPL (Table 2).

Direct comparison of response time courses to spoken and written narratives

Overlap in the reliability of responses across the spoken and written conditions does not tell us whether the response time courses for individual sentences embedded within a real-life narrative are similar across the two conditions. To test for the direct correspondence, we correlated the response time courses in the listeners' brains to the response time courses in the readers' brains within each brain area.

Most linguistic areas and some extra-linguistic areas demonstrated a remarkable invariance to modality, responding similarly to the narrative regardless of whether it was represented aurally or visually (Fig. 5). To illustrate the effect, we first present the mean time course for the spoken and written conditions sampled from two independent ROIs in the left pSTG and precuneus (Fig. 5*A*), followed by whole brain analysis (Fig. 5*B*), and ROI analysis (Fig. 5*C*). Written and spoken narratives clearly evoked similar mean response time courses in the pSTG and precuneus (Fig. 5*A*). Equally modality-invariant responses were observed in the angular gyrus, IFG, anterior dmPFC, and left DLPFC (Fig. 5*B*; Table 2). The results of the voxelwise analysis were reproduced in a group of independent linguistic ROIs (see Materials and Methods), such as the left IFG, the left pSTG, and left and right angular gyrus, and extra-linguistic ROIs such as the precuneus (Fig. 5*C*). These areas exhibited similar responses regardless of presentation modality, but did not respond to the unintelligible conditions in either modality.

The time-locking of auditory and visual stimulus onsets cannot account for the cross-modally shared neural responses. We

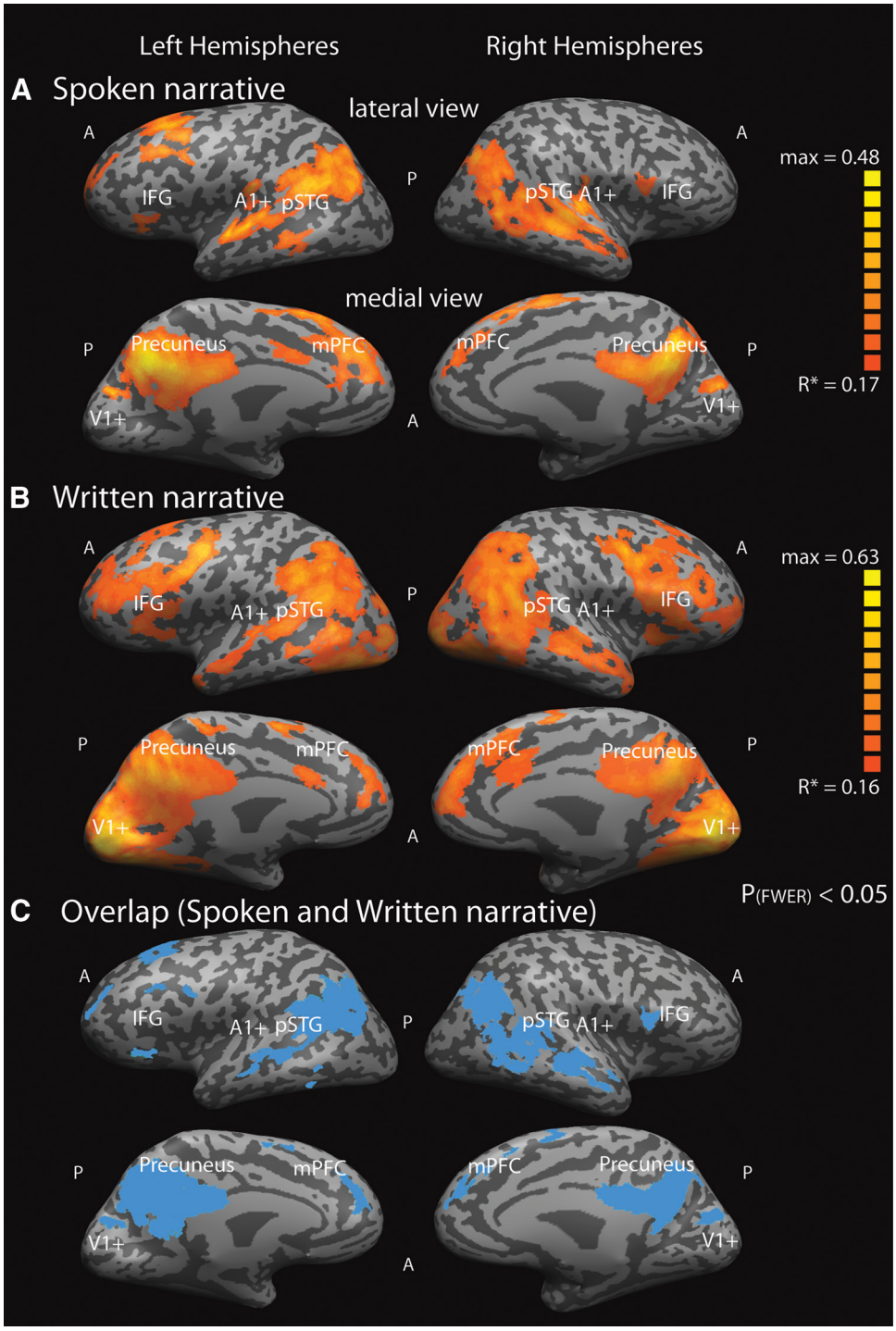

Figure 3. Reliability of brain response within the spoken and the written conditions. The fMRI BOLD time course in each voxel was correlated across subjects to produce a map of inter-SC within each presentation mode. ***A***, ***B***, The surface maps show the areas exhibiting reliable responses for (***A***) subjects who listened to the narrative and (***B***) subjects who read the narrative ($p_{(FEWER)} < 0.05$). ***C***, Brain regions that respond reliably to both the spoken narrative and the written narrative; this is the intersection of the areas shown in ***A*** and ***B***.

tested the magnitude of this low-level onset effect by presenting unintelligible scrambled letters in a time-locked rhythm with the full spoken story. This control stimulus did not exhibit any significant correlations with the full spoken story across the entire brain using the same corrected threshold ($p_{(FEWER)} < 0.05$). In addition, the correlations across subjects within the scrambled-letters condition were reliable only within visual cortex, and not in any of the high-order areas that exhibited modality-invariant responses in other conditions (Figs. 4*B*, 5*C*). These data rule out the possibility that regular stimulus onsets could have elicited the cross-modally reliable responses.

Discussion

In this study, we compared brain responses within and across subjects who either listened to a real-life spoken narrative or read a time-locked presentation of its transcript. Analysis of the temporally extended neural responses revealed two novel findings. The first finding is of robust modality-invariant response time

Table 2. Talairach coordinates of the statistical maps

	Area	*x*	*y*	*z*	S	W	S + W	S > W	W > S
Left hemisphere	pSTG	−57	−51	17	✘	✘	✘		
	aSTG	−58	−24	0	✘	✘	✘		
	Temporal pole	−45	13	−18		✘			✘
	Angular gyrus	−47	−61	18	✘	✘	✘		
	Supramarginal gyrus	−55	−50	21	✘	✘	✘		
	aIPL	−48	−45	46		✘			✘
	pIPL	−48	−59	38	✘	✘	✘		
	Dorsal pars triangularis (IFG)	−42	20	11		✘			✘
	Ventral pars triangularis (IFG)	−41	21	4	✘	✘	✘		
	Orbital cortex	−45	22	1	✘	✘			
	Pars opercularis (IFG)	−45	13	7	✘	✘	✘		
	Precuneus	−3	−59	30	✘	✘	✘		
	Posterior cingulate cortex	−1	−31	24	✘	✘	✘		
	Retrosplenial	−5	−47	8	✘	✘			
	Anterior dmPFC	−7	48	28	✘	✘	✘		
	Posterior dmPFC	−7	41	24	✘	✘		✘	
	Anterior DLPFC	−44	20	37	✘			✘	
	Posterior DLPFC	−44	5	37		✘			✘
Right hemisphere	pSTG	49	−51	9	✘	✘	✘		
	aSTG	50	−26	0	✘	✘	✘		
	Temporal pole	48	10	−17	✘	✘	✘		
	Angular gyrus	41	−66	38	✘	✘	✘		
	Supramarginal gyrus	50	−54	30	✘	✘	✘		
	aIPL	51	−45	45		✘			✘
	pIPL	45	−55	44	✘	✘	✘		
	Dorsal pars triangularis (IFG)	50	23	10		✘			✘
	Ventral pars triangularis (IFG)	50	17	4	✘	✘	✘		
	Orbital cortex	46	36	−6		✘			✘
	Pars opercularis (IFG)	50	15	14	✘	✘	✘		
	Precuneus	3	−55	33	✘	✘	✘		
	Posterior cingulate cortex	1	−34	24	✘	✘	✘		
	Retrosplenial	5	−48	9	✘	✘	✘		
	Anterior dmPFC	6	43	30	✘	✘	✘		
	Posterior dmPFC	9	31	34		✘			✘
	DLPFC	35	5	32		✘			✘

The Talairach coordinates were derived from the following statistical maps: the inter-SC within the spoken condition (S) and the written condition (W; Fig. 3*A*,*B*); modality-selective responses to the spoken condition (S > W) and to the written condition (W > S; Fig. 4*A*, red and green); inter-SC across the spoken and written conditions (S + W; Fig. 5*B*).

a/pSTG, Anterior/posterior superior temporal gyrus; a/pIPL, anterior/posterior inferior parietal lobule; IFG, inferior frontal gyrus; dmPFC, dorsomedial prefrontal cortex; DLPFC, dorsolateral prefrontal cortex.

courses within language-related areas along the pSTG and the IFG, as well as the precuneus (Fig. 5). The invariant response time courses indicate that, not only do these regions process real-life linguistic inputs of multiple modalities, they process that information in a similar fashion across modalities. The second finding is of modality-selective responses, which were not restricted to early visual and auditory cortices, but were also observed in parietal and frontal cortices. Although the sensory regions are expected to exhibit modality-selectivity, the observation of selective responses in parietal and frontal areas is surprising in light of their suggested amodal control functions (Mesulam, 1998; Chee et al., 1999).

Modality-invariant responses

The spoken and written narratives we presented have little in common in terms of their low-level sensory properties. This fit well with our observation of modality-selective response patterns in early visual and auditory areas (Fig. 4*B*). Nevertheless, both forms convey essentially the same meaning as verified by our comprehension test (Fig. 2). The behavioral invariance was paralleled by robust response invariance within language-related areas. Prior studies reported spatial overlap between areas that respond to spoken language and areas that respond to written text (Raij et al., 2000; Shaywitz et al., 2001; Booth et al., 2002; Marinković, 2004; van Atteveldt et al., 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012). In principle, a brain area could respond reliably to both spoken and written narratives, but with one temporal response profile for the spoken narrative and another for the written narrative. This study goes beyond these results by demonstrating that the response time courses for real-life complex narratives across the two conditions were highly similar within the regions noted above (Fig. 5). Moreover, the shared responses across spoken and written language extended to the precuneus, a high-order area whose responses are strongly contextually modulated (Ben-Yakov et al., 2012) and which does not respond reliably to streams of unrelated words or sentences (Xu et al., 2005; Lerner et al., 2011).

The similarity in neural responses across spoken and written stimuli in pSTG, left IFG, and precuneus may arise from the grammatical structures, lexical items, and situation models (van Dijk and Kintsch, 1983; Zwaan and Radvansky, 1998; Fairhall and Caramazza, 2013) that are shared by the spoken and written stimuli. Shared responses in many of the same temporal and parietal areas were also observed across Russian-speakers who listened to a Russian narrative and English-speakers who listened to its English translation (Honey et al., 2012). The partial invariance to both modalities and languages in these areas indicates

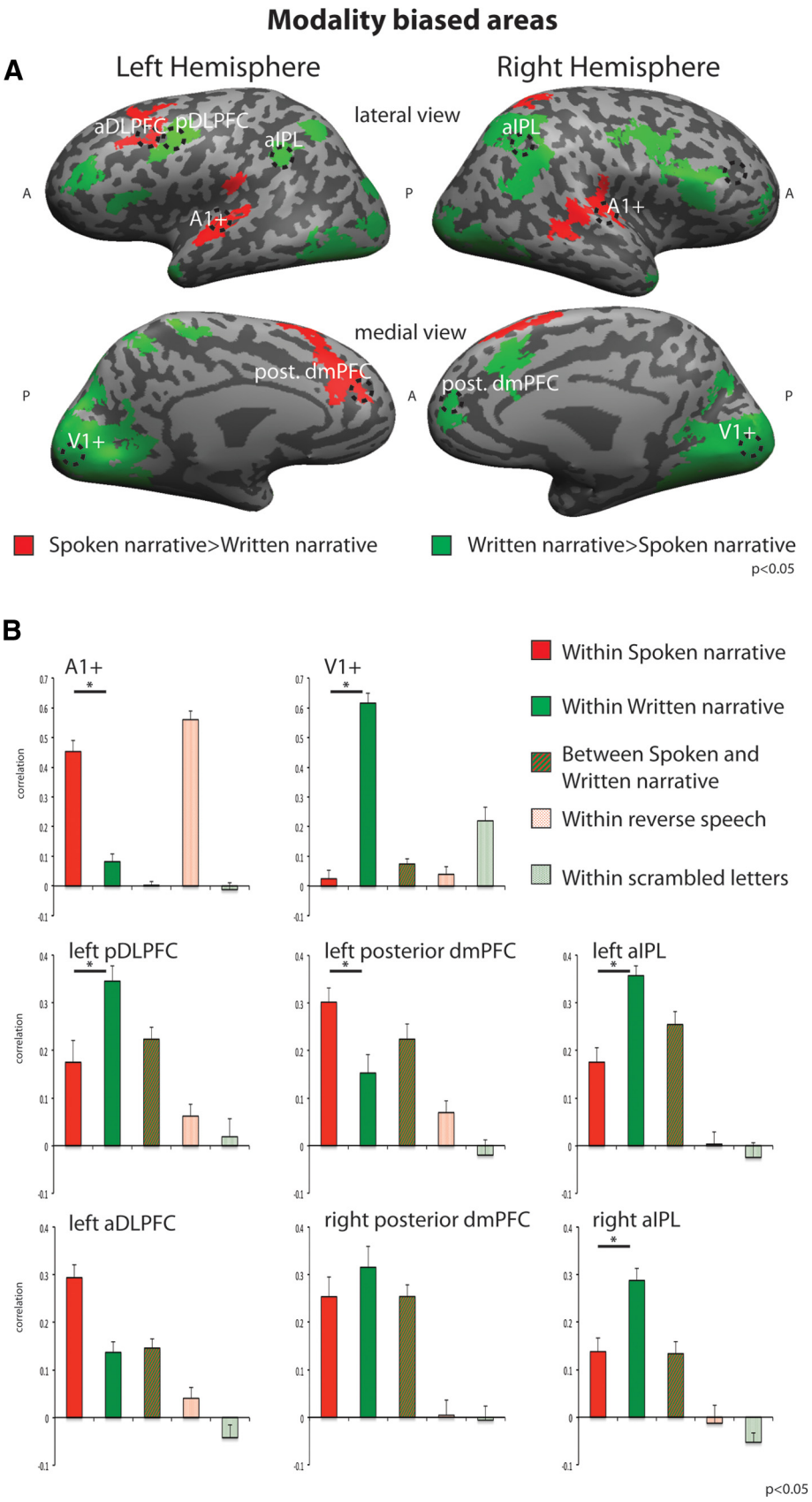

Figure 4. Brain regions exhibiting modality-selective bias to the spoken or written narratives. ***A***, *t* tests between the spoken and written conditions within each voxel revealed areas that exhibit more reliable responses in the spoken condition (red) and other areas that exhibit more reliable responses in the written condition (green). Dashed circles represent ROIs locations. ***B***, Independently defined ROIs that exhibit modality-selective neural responses to spoken and written narratives. These regions include primary auditory area (A1+), primary visual area (V1+), left pDLPFC and aDLPFC, the left and right posterior dmPFC, and anterior regions in the left and right IPL.

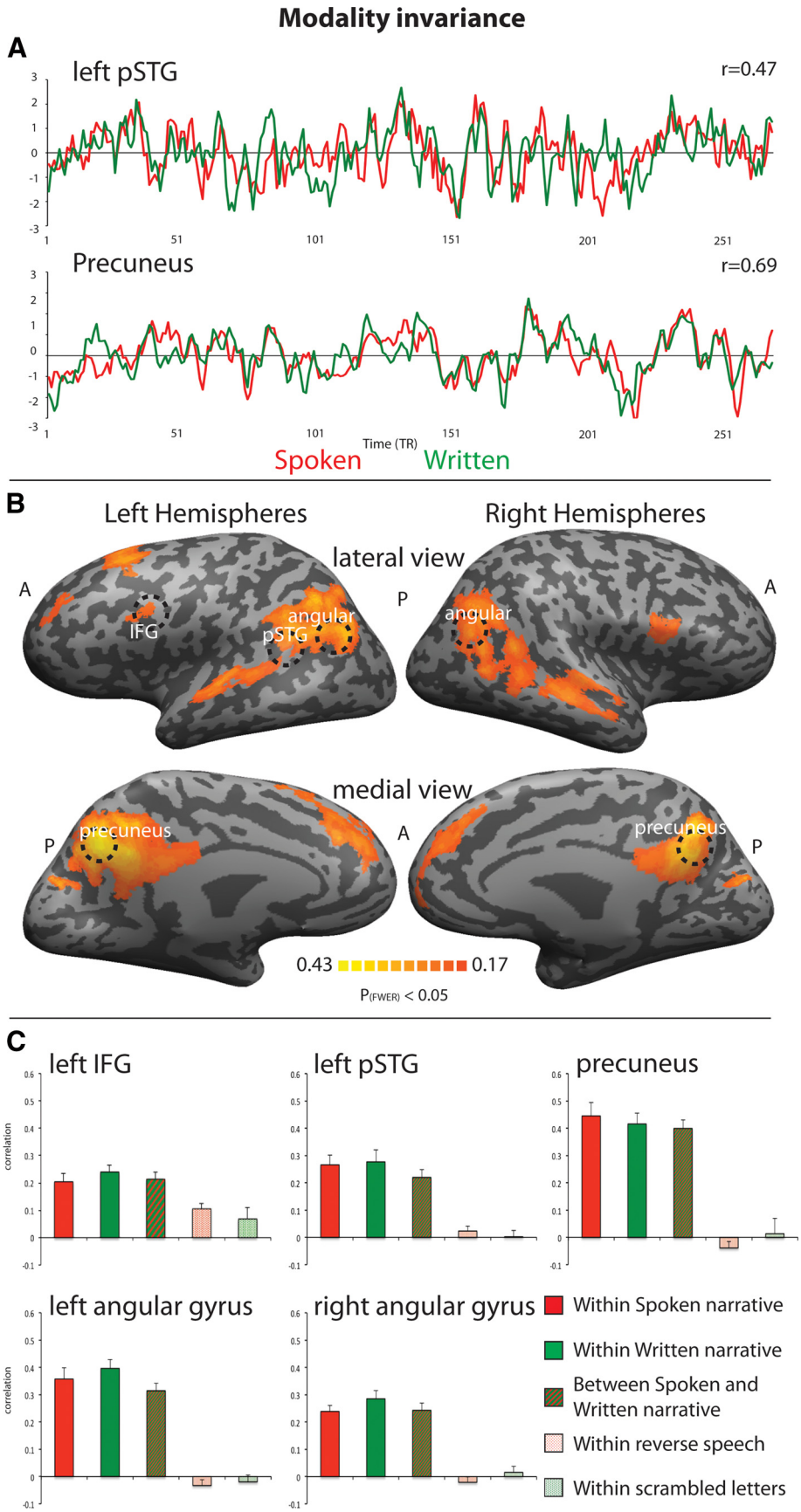

Figure 5. Brain regions exhibiting modality-invariant responses to the spoken and the written narratives. ***A***, The average time courses of the responses in the left pSTG and the precuneus evoked by the written (green) and spoken (red) narrative. ***B***, The fMRI BOLD time course in each area was correlated across conditions to produce a map of inter-SC across modes of presentation. Dashed circles represent ROI locations. ***C***, Independently defined regions of interest that exhibit modality-invariant neural responses to spoken and written narratives. These regions include the left IFG, the left pSTG, the precuneus, and the left and right angular gyrus.

that their representations are highly abstracted from the sensory input.

Given the uncommon reading task, where words appear in the middle of the screen at a fixed rate, our design is not suitable for revealing additional processes (such as the control of eye movements), which are unique to the reading of written text. However, our behavioral results point toward similar levels of comprehension and engagement with our stimuli across the groups. Moreover, such task differences cannot induce similarities in the neural activity across conditions; rather, they will tend to reduce the correlation across subjects who read and listened to the story. Thus, the actual invariant responses across reading and listening may be even more extensive than reported in this study.

The similarity of response time courses to spoken and written narratives attests to the plasticity of the language system. Regions that exhibit modality-invariant responses in the present study would have processed only auditory language signals within the first few years of life, before written language skills were acquired. Remarkably, the human nervous system learns to extract similar information from purely visual signs. In earlier stages of language processing, within the superior temporal cortex, this invariance may reflect the encoding of visual information (graphemes) into originally auditory representations (phonemes; Calvert et al., 2000; Raij et al., 2000). However, in regions further away from the auditory cortex and the STG, the modality-invariant responses more likely reflect amodal information processing elicited in a similar fashion by auditory and visual input.

Modality-selective responses

Surprisingly, a subset of parietal and frontal cortices exhibited strong preference for one modality over the other (Fig. 4). In particular, we observed a greater reliability for spoken narratives in the left anterior DLPFC, and a greater reliability for written narratives in the left posterior DLPFC. Similarly, the responses in the left (right) posterior dmPFC were more reliable for spoken (written) narratives. Finally, responses in the lateral anterior parietal cortex were more reliable for the written version of the narrative. The double-dissociated selective responses such as in the anterior and posterior portions of left DLPFC may indicate differential frontal cortical involvement in the active maintenance of auditory and visual information.

The functional selectivity observed in frontal areas in this study is consistent with nonhuman primate studies that have revealed response selectivity for faces in anterior ventrolateral prefrontal cortex (VLPFC) and for vocalizations in posterior VLPFC (Romanski, 2007). In addition, distinct frontoparietal networks in humans have been associated with memory for auditory and visual inputs (Protzner and McIntosh, 2007) and with attention to frequency-based auditory information and spatial-based visual information (Braga et al., 2013).

The fact that distinct subsections of medial and lateral frontal cortex exhibited a preference for spoken over written language (and vice versa; Fig. 4) suggests that spoken and written language inputs may induce distinct control processes. These distinct control processes may be related to the sensory modality (i.e., audio vs visual) as well as to the task modality (i.e., reading vs listening). At the same time, it appears that the modality-specific information that reaches these frontal and parietal regions is not low-level sensory information, because the higher-order regions only respond reliably to the meaningful linguistic stimuli and not to unintelligible scrambled letters or sounds (Figs. 4, 5).

Although our inter-SC analysis method is successful at characterizing the neural dynamics that are shared over time across spoken and written natural conditions, it also has its limitations. First, more spatially refined methods, such as fMRI-adaptation, are needed to map the neural organization of writing-selective, speech-selective, and amodal neurons at subvoxel resolution (Grill-Spector and Malach, 2001; van Atteveldt et al., 2010). Second, identifying areas with superadditive responses to simultaneous spoken and written stimuli could potentially indicate how neurons in these areas integrate information across modalities (Calvert, 2001).

In conclusion, the present study reveals modality-invariant and modality-selective responses to written and spoken narrative by directly comparing response time courses across listeners and readers. First, we observed that real-life narratives evoked reliable responses across many brain areas, ranging from early sensory areas to linguistic areas, and up to high-order parietal and frontal areas. Second, we observed a remarkable invariance to input form in linguistic areas, which responded similarly to the spoken and written narratives. However, the strong modality-invariance in these linguistic areas was accompanied by modality-selective responses in high-order parietal and frontal cortices. These findings challenge the classical distinction between sensory unimodal areas and higher-order amodal areas by demonstrating that some higher-order areas can display strong invariance to the input modality, whereas other areas can retain strong selectivity.

References

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17–26. CrossRef Medline

Ben-Yakov A, Honey CJ, Lerner Y, Hasson U (2012) Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. Neuroimage 63:501–506. CrossRef Medline

Bestelmeyer PE, Belin P, Grosbras MH (2011) Right temporal TMS impairs voice detection. Curr Biol 21:R838–R839. CrossRef Medline

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. CrossRef Medline

Braga RM, Wilson LR, Sharp DJ, Wise RJ, Leech R (2013) Separable networks for top-down attention to auditory non-spatial and visuospatial modalities. Neuroimage 74:77–86. CrossRef Medline

Brainard DH (1997) The psychophysics toolbox. Spatial Vision 10:433–436. CrossRef Medline

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

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657. CrossRef Medline

Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and visual word processing studied with fMRI. Hum Brain Mapp 7:15–28. CrossRef Medline

Dinstein I, Hasson U, Rubin N, Heeger DJ (2007) Brain areas selective for both observed and executed movements. J Neurophysiol 98:1415–1427. CrossRef Medline

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. J Neurosci 33:10552–10558. CrossRef Medline

Ferstl EC, Neumann J, Bogler C, von Cramon DY (2008) The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Hum Brain Mapp 29:581–593. CrossRef Medline

Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD (1995) Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57:109–128. CrossRef Medline

George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM (1996) Understanding emotional prosody activates right hemisphere regions. Arch Neurol 53:665–670. CrossRef Medline

Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416–429. CrossRef Medline

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. CrossRef Medline
Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640. CrossRef Medline
Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. Trends Cogn Sci 14:40–48. CrossRef Medline
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline
Honey CJ, Thompson CR, Lerner Y, Hasson U (2012) Not lost in translation: neural responses shared across languages. J Neurosci 32:15277–15283. CrossRef Medline
Houston SD (2004) The first writing. Cambridge, United Kingdom: Cambridge UP.
Jobard G, Vigneau M, Mazoyer B, Tzourio-Mazoyer N (2007) Impact of modality and linguistic complexity during reading and listening tasks. Neuroimage 34:784–800. CrossRef Medline
Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci 31:2906–2915. CrossRef Medline
Lindenberg R, Scheef L (2007) Supramodal language comprehension: role of the left temporal lobe for listening and reading. Neuropsychologia 45:2407–2415. CrossRef Medline
Marinković K (2004) Spatiotemporal dynamics of word processing in the human cortex. Neuroscientist 10:142–152. CrossRef Medline
Mesulam MM (1998) From sensation to cognition. Brain 121:1013–1052. CrossRef Medline
Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25. CrossRef Medline
Pelli DG (1997) The VideoToolbox software for visual psychophysics transforming numbers into movies. Spat Vis 10:437–442. CrossRef Medline
Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci 1191:62–88. CrossRef Medline
Protzner AB, McIntosh AR (2007) The interplay of stimulus modality and response latency in neural network organization for simple working memory tasks. J Neurosci 27:3187–3197. CrossRef Medline
Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28:617–625. CrossRef Medline
Romanski LM (2007) Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. Cereb Cortex 17:i61–i69. CrossRef Medline
Romanski LM, Averbeck BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32:315–346. CrossRef Medline
Ross ED, Mesulam MM (1979) Dominant language functions of the right hemisphere? Prosody and emotional gesturing. Arch Neurol 36:144–148. CrossRef Medline
Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science 326:445–449. CrossRef Medline
Shaywitz BA, Shaywitz SE, Pugh KR, Fulbright RK, Skudlarski P, Mencl WE, Constable RT, Marchione KE, Fletcher JM, Klorman R, Lacadie C, Gore JC (2001) The functional neural architecture of components of attention in language-processing tasks. Neuroimage 13:601–612. CrossRef Medline
Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ (2006) Converging language streams in the human temporal lobe. J Neurosci 26:7328–7336. CrossRef Medline
Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.
Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A temporal bottleneck in the language comprehension network. J Neurosci 32:9089–9102. CrossRef Medline
van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. Neuron 43:271–282. CrossRef Medline
van Atteveldt NM, Blau VC, Blomert L, Goebel R (2010) fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci 11:11. CrossRef Medline
van Dijk T, Kintsch W (1983) Strategies of discourse comprehension. New York: Academic.
Xu J, Kemeny S, Park G, Frattali C, Braun A (2005) Language in context: emergent features of word, sentence, and narrative comprehension. Neuroimage 25:1002–1015. CrossRef Medline
Zarahn E, Aguirre GK, D'Esposito M (1997) Empirical analyses of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage 5:179–197. CrossRef Medline
Zwaan RA, Radvansky GA (1998) Situation models in language comprehension and memory. Psychol Bull 123:162–185. CrossRef Medline