Convergence of spoken and written language processing in the superior temporal sulcus

Stephen M. Wilson [a,*], Alexa Bautista [b], Angelica McCarron [b]

[a] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA
[b] Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ, USA

ABSTRACT

Spoken and written language processing streams converge in the superior temporal sulcus (STS), but the functional and anatomical nature of this convergence is not clear. We used functional MRI to quantify neural responses to spoken and written language, along with unintelligible stimuli in each modality, and employed several strategies to segregate activations on the dorsal and ventral banks of the STS. We found that intelligible and unintelligible inputs in both modalities activated the dorsal bank of the STS. The posterior dorsal bank was able to discriminate between modalities based on distributed patterns of activity, pointing to a role in encoding of phonological and orthographic word forms. The anterior dorsal bank was agnostic to input modality, suggesting that this region represents abstract lexical nodes. In the ventral bank of the STS, responses to unintelligible inputs in both modalities were attenuated, while intelligible inputs continued to drive activation, indicative of higher level semantic and syntactic processing. Our results suggest that the processing of spoken and written language converges on the posterior dorsal bank of the STS, which is the first of a heterogeneous set of language regions within the STS, with distinct functions spanning a broad range of linguistic processes.

Introduction

Spoken and written language take very different perceptual forms. The speech waveform enters the auditory system as a continuous stream containing spectro-temporal cues to phonemes that the listener must segment and map onto phonological word forms. In contrast, written language enters the brain in the form of patterns of light on the retina; the reader makes saccades to fixate on successive chunks of text, identifies letters, and maps them onto orthographic word forms. In either case, the final goal is the same: to derive a conceptual representation of meaning. But to get to that endpoint, there are also processing stages that are largely independent of the input modality, for instance, accessing the meanings of words from their forms, combining their meanings according to the syntactic structure of the utterance, and so on. These basic observations suggest a "Y-shaped" model of spoken and written language processing, in which two distinct modality-specific streams of processing converge at some point onto a modality-neutral common processing stream, which ultimately yields an abstract representation of meaning.

The cortical pathways involved in the early, modality-specific stages of processing of both spoken and written language are quite well understood. For spoken language, primary and higher level auditory areas in Heschl's gyrus and on the dorsal and lateral surfaces of the superior temporal gyrus (STG) carry out spectro-temporal analysis of the auditory signal (Binder et al., 1996; Formisano et al., 2003; Mesgarani et al., 2014; see Moerel et al., 2014 for review). For written language, a hierarchy of occipital and ventral temporal regions in the ventral visual stream code increasingly complex and abstract visual features of the letter string (Binder and Mohr, 1992; Cohen et al., 2000; Vinckier et al., 2007; Dehaene and Cohen, 2011). The cortical correlates of the conceptual representations that constitute the endpoint of language comprehension are also increasingly well understood. This semantic system comprises a network of brain regions including the middle temporal gyrus (MTG), anterior temporal lobe, angular gyrus, and inferior frontal gyrus (IFG) (Geschwind, 1965; Patterson et al., 2007; Binder et al., 2009; Visser et al., 2012; Huth et al., 2016).

What is less clear is the functional neuroanatomy of the intervening processes and representations, including precisely how and where the processing of spoken and written language converges. Several functional imaging studies have shown that neural activity common to the processing of spoken and written language is localized to the superior temporal sulcus (STS), predominantly in the left hemisphere (Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Berl et al.,

2010). Moreover, the STS is similarly modulated by rate and intelligibility in both modalities (Vagharchakian et al., 2012), and the time courses of STS responses to the same linguistic material in spoken and written form are remarkably similar (Regev et al., 2013). Taken together, these studies suggest that spoken and written language processing converge in the STS.

While this finding is a vital first step, it leaves many important questions unanswered, because the STS is a not a unitary structure (Liebenthal et al., 2014). Rather, it is a deep sulcus containing a great expanse of neural tissue. Studies in non-human primates have shown that the STS contains numerous subdivisions with distinct cytoarchitectonic properties and connectivity profiles (Jones and Powell, 1970; Seltzer and Pandya, 1978). In the domain of language, the STS has been implicated in a heterogeneous range of processes, covering the gamut of stages from sublexical processing of speech (Liebenthal et al., 2005; Möttönen et al., 2006; Uppenkamp et al., 2006; Turkeltaub and Coslett, 2010; Liebenthal et al., 2014), to representation of phonological word forms (Okada and Hickok, 2006), to semantic and syntactic processing (Scott et al., 2000; Davis and Johnsrude, 2003; Friederici et al., 2009; Wilson et al., 2016). In both the spoken and written modalities, regions in the STS are sensitive to manipulation of lower level (van Atteveldt et al., 2004) and higher level (Xu et al., 2005; Jobard et al., 2007) aspects of language processing.

To better understand how spoken and written language processing streams converge in the STS, it is first necessary to clarify the functional parcellation of the STS with respect to language. This undertaking faces two main challenges: one linguistic, and the other anatomical. The first challenge is that language processing generally involves seamless and integrated computations at multiple levels: phonological or orthographic, lexical, semantic, syntactic and so on. In functional imaging studies, even the most ingenious contrasts between conditions (e.g. Scott et al., 2000) often end up entailing multiple differences between conditions, at more than one level of representation (Binder, 2000). In the present study, we addressed this challenge by investigating not only contrasts between carefully matched intelligible and unintelligible spoken and written inputs, but also by quantifying neural responses to the unintelligible inputs themselves (Woodhead et al., 2011). Models of spoken and written language processing (e.g. McClelland and Rumelhart, 1981; McClelland and Elman, 1986; Taylor et al., 2013) make clear predictions about the extent to which different kinds of unintelligible inputs should drive different levels of linguistic processing. Furthermore, we used searchlight multi-voxel pattern analysis (MVPA; Kriegeskorte et al., 2006) to identify brain regions that can distinguish between different inputs by means of distributed patterns of signal change, even if they show the same overall level of activation (Haxby et al., 2001; Kamitani and Tong, 2005).

The second challenge to parcellating the STS is anatomical: the dorsal and ventral banks of the STS are, by nature, in close physical proximity to one another, and functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) generally lack the spatial resolution to distinguish between activity on the two banks of the sulcus. While fMRI has higher spatial resolution than PET, the blood oxygen level-dependent (BOLD) signal that is the basis of most fMRI studies is more sensitive to signal changes in draining veins than in gray matter itself (Bandettini and Wong, 1997; Lai et al., 1993; Menon et al., 1993), and medium-sized draining veins run through the STS, as they do through all major sulci. Therefore, in typical fMRI studies, activations in the STS are localized to the veins that run through the sulcus, which are downstream of the location(s) where neural activity is occurring, and are therefore somewhat uninformative with regard to the specific site of the neural activity (Wilson, 2014). To address this challenge, we employed several strategies to maximize spatial resolution. First, small voxels were acquired, and no spatial smoothing was applied. Second, a breath-holding task in a separate run was used to estimate and correct for voxelwise differences in cerebrovascular reactivity (CVR) (Bandettini and Wong, 1997; Cohen et al., 2004; Handwerker et al., 2007; Thomason et al., 2007; Murphy et al., 2011; Wilson, 2014); this effectively de-emphasizes

signal from veins, which have very high CVR (Wilson, 2014). Third, veins were identified on susceptibility-weighted imaging (SWI), and masked out. Fourth, intersubject normalization was carried out with the large-deformation DARTEL registration algorithm (Ashburner, 2007), which aligns specific structures across participants with exceptional accuracy (Klein et al., 2009). Taken together, these methodological choices were intended to facilitate the identification of distinct patterns of responses to intelligible and unintelligible spoken and written inputs on the dorsal and ventral banks of the STS, in order to further our understanding of how spoken and written language processing streams converge in the STS.

Materials and methods

Participants

Sixteen healthy participants of a wide range of ages took part in the study (mean age = 57 years; range = 21–81 years; 9 females; 1 left-hander and 2 ambidextrous). No participant reported any history of neurological disorders. All participants gave written informed consent, and the study was approved by the institutional review board at the University of Arizona.

Narrative comprehension paradigm

Each participant completed two ($N = 5$) or three ($N = 11$) narrative comprehension runs. There were five conditions: listening to spoken narrative segments ("Spoken"), listening to backwards spoken narrative segments ("Backwards"), reading written narrative segments ("Written"), quasi-reading scrambled written narrative segments ("Scrambled"), and no stimulus ("Rest"). Each run comprised 15 segments per condition, presented in pseudorandom order. A sparse sampling protocol was used, with a repetition time (TR) of 9500 ms and an acquisition time (TA) of 2269 ms, leaving 7231 ms silence between successive acquisitions. Two initial volumes were acquired and discarded, and then one image was acquired after each stimulus or rest period, for a total of 75 volumes per run.

The narrative was the beginning of an audiobook recording of the novel *Hope Was Here* by Joan Bauer, read by Jenna Lamua (Bauer, 2004). The narrative was split into segments at pauses such that each segment was as long as possible up to 7 s (occasionally, slightly longer segments were extracted, then reduced to 7 s by shortening internal pauses). The mean length of the segments was 5656 ms ± 1012 (SD) ms.

In the Spoken narrative condition (Fig. 1A), each narrative segment was presented centered in the silent interval between scans, such that the peak of a typical hemodynamic response to the segment would coincide with acquisition of the subsequent image.

The Backwards narrative condition (Fig. 1B) was the same, except that the segments were played in reverse, rendering them unintelligible. Note that backwards speech contains partial phonemic information. In particular, monophthongal vowels are not greatly affected by reversal, and many consonants also retain their identities. Previous research has shown that naive transcription of backwards words is considerably better than chance (Binder et al., 2000), supporting the notion that backwards speech carries phonemic information; it seems plausible that phonemic information could also be extracted from backwards sentences. Models of spoken word comprehension generally posit that representations of phonemes are mapped onto lexical modes by a spreading activation mechanism (McClelland and Elman, 1986). From this perspective, because it contains recognizable phonemes, the Backwards condition would be expected to activate brain regions involved in phonemic representation of spoken inputs. Moreover, due to spreading activation between phonemic and lexical representations, the Backwards condition should also activate brain regions involved in representation of lexical nodes, even though no lexical nodes will ultimately be selected. Because no lexical nodes are selected, brain regions involved in semantic
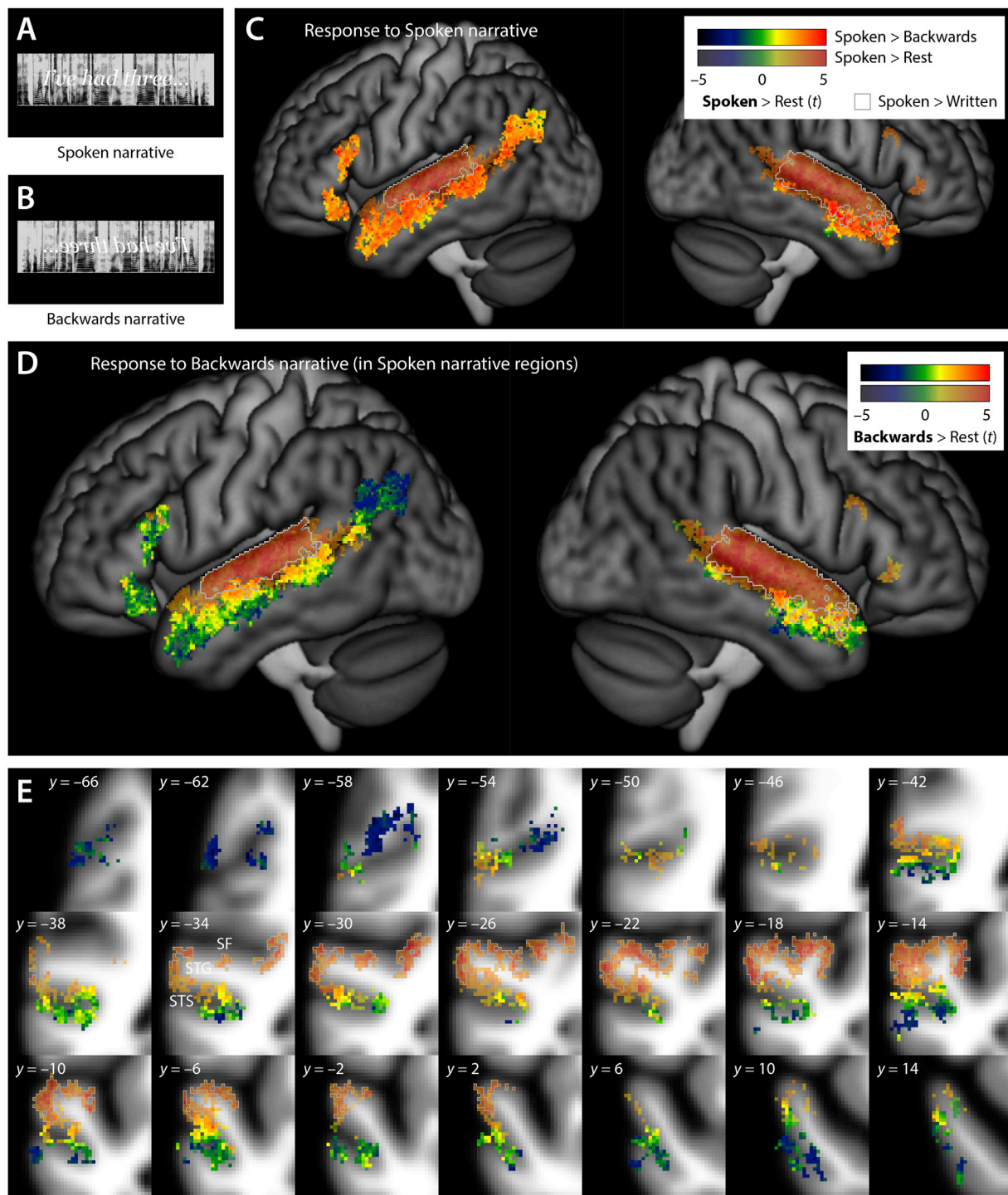
Fig. 1. Spoken language comprehension. (A) The Spoken condition involved segments of spoken narrative speech. (B) In the Backwards condition, these segments were reversed and therefore unintelligible. (C) Brain regions involved in any aspect of spoken language comprehension were identified by plotting the union of two contrasts: Spoken – Rest, shown in partially transparent color, and Spoken – Backwards, shown in opaque color. For the regions activated by either of these contrasts, the colors indicate the *t* statistic for the contrast of Spoken – Rest. Regions activated for the contrast of Spoken – Written (i.e. likely auditory areas) are shown with gray outlines. (D) This panel shows how the brain regions identified in panel (C) as involved in any aspect of spoken language comprehension responded to unintelligible spoken language: the color indicates the *t* statistic for the contrast of Backwards – Rest. Voxels are shown partially transparent or opaque depending on whether they were activated for Spoken – Rest or Spoken – Backwards respectively, just as in the previous panel. (E) The same data as in (D), but shown on a series of coronal slices 4 mm apart through the left STS.

representations, or any higher level processes, should not be activated. This supposition is supported by priming studies, which have shown that nonwords do not result in semantic priming unless they are very similar to real words (Connine et al., 1993). This suggests that nonwords do not activate semantic representations, therefore neither should backwards

speech.

The Written narrative condition (Fig. 2A) was created by transcribing the words in each segment along with their exact timing. In written narrative segments, the words of the narrative were presented one at a time with the same timing as the spoken narrative. The words were
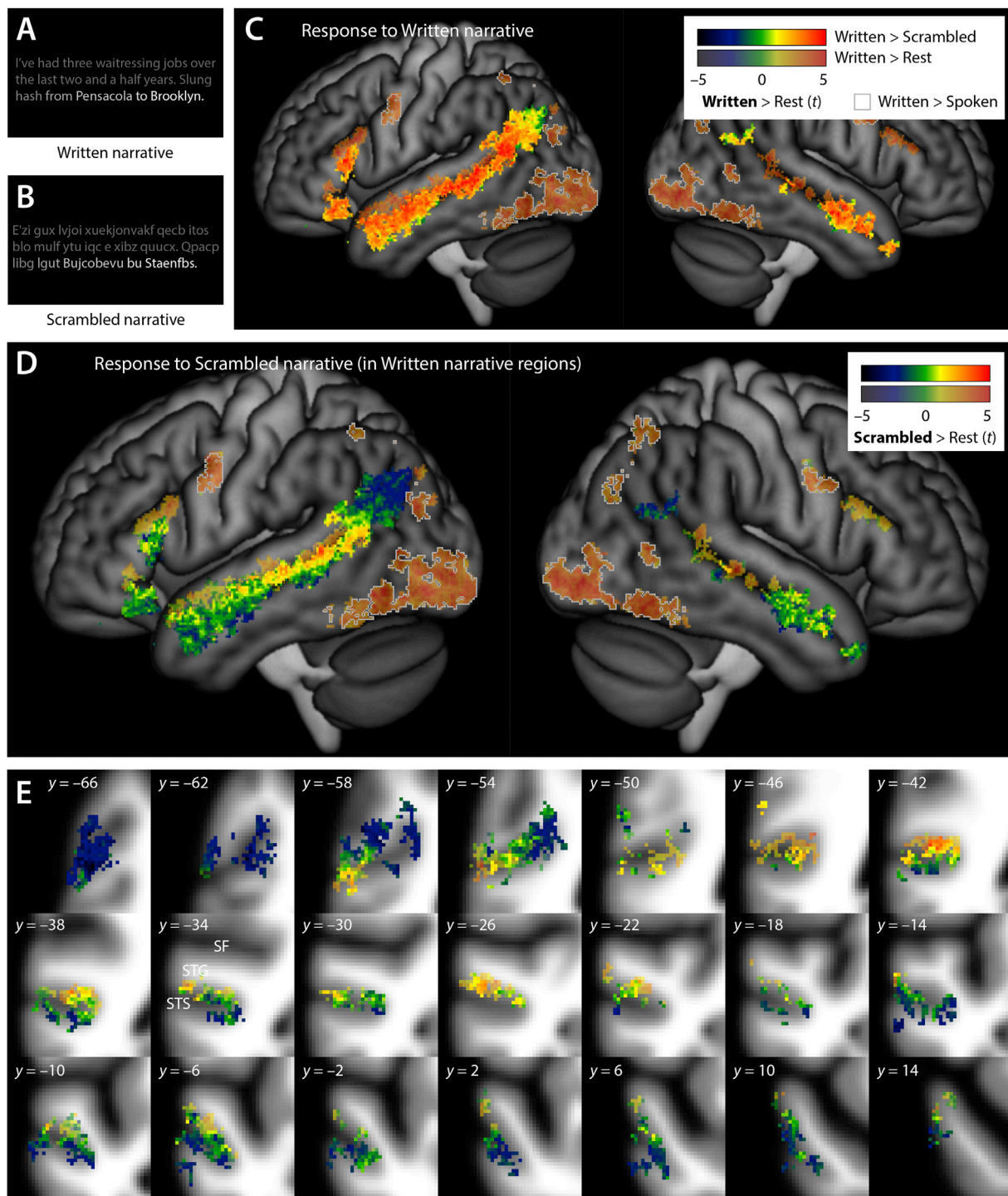
Fig. 2. Written language comprehension. (A) The Written condition involved segments of written narrative text. (B) In the Scrambled condition, consonants and vowels were randomly replaced with other consonants and vowels, rendering the text unintelligible. (C) Brain regions involved in any aspect of written language comprehension were identified by plotting the union of two contrasts: Written – Rest, shown in partially transparent color, and Written – Scrambled, shown in opaque color. For the regions activated by either of these contrasts, the colors indicate the *t* statistic for the contrast of Written – Rest. Regions activated by the contrast of Written – Spoken (i.e. likely auditory areas) are shown with gray outlines. (D) This panel shows how the brain regions identified in panel (C) as involved in any aspect of written language comprehension responded to unintelligible written language: the color indicates the *t* statistic for the contrast of Scrambled – Rest. Voxels are shown partially transparent or opaque depending on whether they were activated for Written – Rest or Written – Scrambled respectively, just as in the previous panel. (E) The same data as in (D), but shown on a series of coronal slices 4 mm apart through the left STS.

presented from left to right, with previously presented words remaining on the screen (but fading slightly over time), such that by the end of the segment, all of the text was on the screen. Some segments fit on one line, while others required two lines. This mode of stimulus presentation

exactly matched the timing of the spoken narrative condition, while providing a more natural reading experience than rapid serial visual presentation paradigms, in which each word replaces the previous one. Similar approaches have been employed in several recent studies (Hillen

et al., 2013; Choi et al., 2014).

The Scrambled narrative condition (Fig. 2B) was similar, except that every consonant was replaced with a different random consonant (with equal probability), and every vowel was replaced with a different random vowel (with equal probability), rendering the segments unintelligible. This procedure yielded 10.9% real words (many of very low frequency), 43.2% nonwords with one or more orthographic neighbors (median 3 neighbors, mean 5.5 ± 4.9 neighbors) and 45.9% nonwords with no orthographic neighbors. Models of written word comprehension generally posit that representations of graphemes are mapped onto lexical nodes by a spreading activation mechanism (McClelland and Rumelhart, 1981). By the same reasoning described above, the Scrambled narrative condition should activate brain regions involved in orthographic representation of written inputs and those involved in representation of lexical nodes (because of spreading activation). In support of this assumption, a simulation study has shown partial activation of lexical representations by nonwords (Taylor et al., 2013), though it must be noted that the nonwords in that study were matched to real words for orthographic neighborhood size, and thus were more word-like than the nonwords in the present study. The Scrambled narrative condition should not activate brain regions involved in semantic or higher level processing. This assumption is supported by a priming study that showed that written nonwords do not activate semantic representations unless they are very similar to real words (Perea and Lupker, 2003), which most of the words in the present study were not.

Taken together, the Spoken and Written segments presented the audiobook in correct temporal progression. The Backwards and Scrambled segments intervened between segments belonging to the two intelligible conditions, but did not disrupt their overall temporal order.

Participants were familiarized with the stimuli before entering the scanner. They were instructed to listen to the spoken narrative and the "strange sounds", and to read the written narrative and the scrambled words. They were explicitly asked to make saccades to the scrambled words as if they were reading. After the scanning session, each participant confirmed that they had heard, read, and comprehended the narrative.

Auditory stimuli were presented using insert earphones (S14, Sensimetrics, Malden, MA) padded with foam to attenuate scanner noise and reduce head movement. The audio volume was adjusted to a comfortable level for each participant. Visual stimuli were presented on a 24″ MRI-compatible LCD monitor (BOLDscreen, Cambridge Research Systems, Rochester, UK) positioned at the end of the bore, which participants viewed through a mirror mounted to the head coil. Auditory and visual stimuli were controlled with the Psychophysics Toolbox version 3.0.10 (Brainard, 1997; Pelli, 1997) running under MATLAB R2012b (Mathworks, Natick, MA) on a Lenovo S30 workstation.

Hypercapnia paradigm

A breath-holding task was used to quantify and adjust for differences between voxels in their capacity to mount a BOLD response (Bandettini and Wong, 1997; Cohen et al., 2004; Handwerker et al., 2007; Murphy et al., 2011; Thomason et al., 2007; Wilson, 2014). The task has been described in detail previously (Wilson, 2014). In brief, breath-holds and paced breathing between breath-holds were cued by a visual display in which a ball moved up and down along a waveform. There were 6 post-exhalation breath-holds, each 13.8 s in length and separated by 27.6 s of paced breathing with a period of 4.6 s.

Neuroimaging protocol

Images were acquired on a Siemens Skyra 3 Tesla scanner with a 32-channel head coil at the University of Arizona. In each narrative comprehension run, 75 T2*-weighted BOLD echo-planar images (plus 2 initial volumes that were discarded) were acquired with the following parameters: 34 axial slices in ascending order; slice thickness = 2 mm plus 0.4 mm gap; field of view = 224×212 mm; matrix = 112×112 interpolated with zero padding to 224×212; TR = 9500 ms; acquisition time (TA) = 2269 ms; TE = 30 ms; flip angle = 90°; GRAPPA acceleration factor = 2; acquired voxel size = $2.0 \times 1.9 \times 2.0$ mm; reconstructed voxel size = $1 \times 1 \times 2$ mm. The field of view included all of the temporal and occipital lobes, and much of the frontal and parietal lobes. The superior parts of the frontal and parietal lobes were not included, nor was the cerebellum.

For the hypercapnia run, acquisition parameters were the same as for the narrative comprehension runs, except that 128 volumes were acquired (plus 2 initial volumes that were discarded) and the TR was 2300 ms; there were no silent gaps between volumes.

To identify veins, an SWI image was acquired with the following parameters: 80 axial slices; slice thickness = 1.2 mm; field of view = 220×192.5 mm; matrix = 384×336; TR = 28 ms; TE = 20 ms; flip angle = 15°; GRAPPA acceleration factor = 2; voxel size = $0.57 \times 0.57 \times 1.20$ mm.

For anatomical reference, a T1-weighted 3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence was acquired with the following parameters: 160 sagittal slices; slice thickness = 0.9 mm; field of view = 240×240 mm; matrix = 256×256; repetition time (TR) = 2300 ms; echo time (TE) = 2.98 ms; flip angle = 9°; GRAPPA acceleration factor = 2; voxel size = $0.94 \times 0.90 \times 0.94$ mm.

Analysis of functional imaging data

The narrative comprehension data were first preprocessed with tools from AFNI version 2016-08-31 (Cox, 1996). Head motion was corrected, with 6 translation and rotation parameters saved for use as covariates, then the data were detrended with a Legendre polynomial of degree 2. No smoothing was carried out.

A general linear model was fit to each narrative run in native space with the program *fmrilm* from the FMRISTAT package (Worsley et al., 2002). Explanatory variables were created for each of the 4 conditions (i.e., Spoken, Backwards, Written, and Scrambled), while the Rest condition formed an implicit baseline. No hemodynamic response function was modeled; instead, each volume was modeled as reflecting the BOLD response to neural activity in response to the immediately preceding segment. The six motion parameters were included as covariates, as were time-series from white matter and CSF regions, and 3 cubic spline temporal trends.

Each narrative run was co-registered to the breath-holding run, then the breath-holding run was co-registered to the T1-MPRAGE anatomical image, using SPM5 (Friston, 2007). The anatomical images were initially normalized to Montreal Neurological Institute (MNI) space using the Unified Segmentation procedure (Ashburner and Friston, 2005) implemented in SPM5 (Friston, 2007), running under MATLAB R2011a. More anatomically precise intersubject registration was then performed with the DARTEL toolbox (Ashburner, 2007) by warping each subject's image to a template created from 50 separate healthy control participants (Wilson et al., 2010a). These transformations were applied to the statistical maps, which were written with $1 \times 1 \times 1$ mm voxels.

The primary contrasts of interest and their interpretations are outlined in Table 1. More detailed information about the interpretation of each contrast is provided in the Results and Discussion sections. For each participant, contrasts derived from the two or three narrative runs were combined in fixed effects models using the FMRISTAT program *multistat*.

The breath-holding data were analyzed as described previously (Wilson, 2014). Each contrast image for each participant was then corrected for voxelwise differences in cerebrovascular reactivity by dividing the image by that participant's voxelwise percent signal change in response to breath-holding. Voxels where the response to breath-holding was less than 0.5% (generally white matter or CSF) were zeroed out. Voxels containing veins on SWI were also masked out.

Random effects group analyses were then carried out based on these CVR-corrected contrast images, using *t*-tests implemented in *multistat*.

Table 1
Primary contrasts of interest.

Contrast	Interpretation	Figure
(Spoken – Rest) *or* (Spoken – Backwards)	any aspect of spoken language comprehension	1C (transparent + opaque)
Spoken – Backwards	higher-level lexical, semantic and syntactic processing of spoken language	1C (opaque)
Backwards – Rest	spectro-temporal processing, phonological representations, spreading activation to lexical nodes	1D,E
Spoken – Written	spectro-temporal processing	1D,E, gray outlines
(Written – Rest) *or* (Written – Scrambled)	any aspect of written language comprehension	2C (transparent + opaque)
Written – Scrambled	higher-level lexical, semantic and syntactic processing of written language	2C (opaque)
Scrambled – Rest	visual processing and saccades, orthographic representations, spreading activation to lexical nodes	2D,E
Written – Spoken	visual processing and saccades	2D,E, gray outlines
(Spoken – Backwards) – (Written – Scrambled)	differential recruitment for higher-level processing of spoken language	
(Written – Scrambled) – (Spoken – Backwards)	differential recruitment for higher-level processing of written language	
Spoken ≠ Written (MVPA)	sensitivity to input modality	3

The voxelwise threshold was set at a relatively liberal value of $p < .01$, due to the fact that no smoothing had been applied. One-tailed tests were used for contrasts where only direction was computed (e.g. Backwards – Rest), and two-tailed tests were used when both directions were of interest (e.g. Spoken – Written and Written – Spoken). Correction for multiple comparisons was carried out based on spatial extent of clusters at $p < .05$, with reference to the null distribution of the largest cluster in 1000 random permutations in which the signs of individual participants' effect size images were randomized (Nichols and Holmes, 2002).

For a follow-up region of interest (ROI) analysis, anatomical ROIs were drawn on the dorsal and ventral banks of the left STS, excluding the fundus, between a posterior extent of MNI $y = -47$ and an anterior extent of MNI $y = 5$.

Multi-voxel pattern analysis

Searchlight multi-voxel pattern analysis (MVPA) was used to identify any brain regions that could differentiate between conditions of interest (specifically, Spoken versus Written) based on distributed patterns of activity. In this approach, a spherical "searchlight" is centered on each voxel in turn, and a multivariate classifier is trained and tested to determine whether the region within the searchlight is capable of discriminating between the conditions of interest (Kriegeskorte et al., 2006).

First, the preprocessed functional time-series images were transformed into MNI space using DARTEL as described above, and written with $2 \times 2 \times 2$ mm voxels. Then, each run was fit with a general linear model such that each successive set of three volumes belonging to the same condition was modeled as a single "example" to be classified. Since there were 15 volumes per condition per run, this meant that there were 5 examples per condition per run, and therefore 10 or 15 examples per condition in total, depending on whether the participant completed two or three runs. The reason that examples were based on three volumes was to strike a balance between maximizing signal-to-noise in each example, and deriving enough examples to train classifiers.

The searchlight analysis was carried out with the *searchmight* toolbox (version: June 23, 2010) (Pereira and Botvinick, 2011). The searchlight comprised a sphere (this required customization) with a radius of 6 mm (i.e. 123 voxels, except at the edges of the brain). Classifiers were trained to discriminate between Spoken and Written examples. The data were split into five folds. One fold at a time was held out. A multivariate linear discriminant analysis (LDA) classifier with a shrinkage estimator for the covariance matrix (Schäfer & Strimmer, 2005) was trained on the four remaining folds, then classification accuracy was tested on the fifth fold. The LDA with shrinkage classifier was selected since it was found to perform well in most circumstances, and was recommended by the authors of the toolbox (Pereira and Botvinick, 2011).

Then, a parallel analysis was carried out using the a univariate LDA classifier trained only on the mean signal in the searchlight (the shrinkage estimator was not applicable since there was only one dimension so no covariance matrix). The increase in accuracy for the multivariate classifier that had access to patterns as well as overall signal,

over the univariate classifier that had access only to overall signal, was calculated for each voxel. This approach was taken so that MVPA results would reflect regions that discriminate between conditions based on patterns of activity, *above and beyond* their ability to differentiate based on overall mean activity. This was the desired outcome, because discrimination based on overall activity was already captured by the standard mass univariate contrasts described above, which offer higher spatial resolution, since MVPA tends to exaggerate the spatial extent of informative regions (Stelzer et al., 2014).

A group analysis was carried out by comparing the accuracy of the multivariate and univariate classifiers with 1-tailed *t*-tests using *multistat*. This analysis was confined to regions involved in language processing that were not already demonstrated in the univariate analysis to respond preferentially to one modality over the other. This was defined as (Spoken – Rest) or (Spoken – Backwards) or (Written – Rest) or (Written – Scrambled), but not any voxel within 6 mm of any activation for (Spoken – Written) or (Written – Spoken). The image was thresholded at voxelwise $p < .01$, then corrected for multiple comparisons using permutation testing as described above.

Results

Spoken language processing

Brain regions involved in any aspect of spoken language comprehension were identified by plotting the union of two contrasts: Spoken – Rest (Fig. 1C, transparent), and Spoken – Backwards (Fig. 1C, opaque). Consistent with much prior research, the Spoken – Rest contrast activated bilateral regions in the STG, and the Spoken – Backwards contrast activated bilateral regions in the anterior STS and MTG, and left-lateralized regions in the posterior STS, angular gyrus (specifically, both banks of the central branch of the caudal STS, dubbed cSTS2 by Segal and Petrides, 2012), and IFG (specifically, the pars orbitalis, and the ascending ramus of the Sylvian fissure, which separates the pars opercularis from the pars triangularis).

We then plotted the response to the Backwards – Rest contrast in the broad set of regions just identified. Lateral projections provide an overview in Fig. 1D, and a series of coronal slices through the left STS shows the details of patterns in Fig. 1E.

A clear dorsal-to-ventral gradient of processing was apparent. Not surprisingly, there was a strong bilateral response to backwards speech in Heschl's gyrus, and on the dorsal and lateral surfaces of the STG (Fig. 1D and E, uniform orange and red in transparent regions). These regions are involved in spectro-temporal processing of the auditory signal, as evidenced by their activation for Spoken – Rest but not Spoken – Backwards. Moreover, they were activated for the contrast of Spoken – Written (Fig. 1D and E, gray outlines).

The dorsal bank of the STS responded to backwards speech. Importantly, dorsal bank regions were not activated by the Spoken – Written contrast (and were activated by both written conditions; see below), so their response to backwards speech cannot be explained in terms of auditory processing. Instead, dorsal bank responses to backwards speech

may reflect processing of phonological representations, phonological word forms, and/or spreading activation to abstract lexical nodes, since these are the levels of spoken word processing that would be expected to be invoked by backwards speech, according to models of spoken word recognition. Most, but not all, dorsal bank regions responded even more strongly to intelligible speech (i.e. they were significantly activated by the Spoken – Backwards contrast).

The response profile generally changed around the fundus of the STS. In sharp contrast to the dorsal bank, there was little to no response to backwards speech in the ventral bank of the sulcus (green). Indeed, in more anterior ventral bank regions, as well as the anterior MTG and the angular gyrus, the backwards speech condition resulted in deactivation relative to rest (blue). Inferior frontal regions also showed little response to backwards speech, with the exception of the most dorsal parts of each of the two activations. The lack of response to backwards speech suggests that all of these regions are involved in higher level linguistic processing, and these regions were generally activated by the Spoken – Backwards contrast.

To confirm the functional distinction between the dorsal and ventral banks of the STS, we examined the mean signal by condition in the dorsal and ventral bank anatomical ROIs. The ventral bank ROI showed no response to the Backwards condition (mean signal change $= -1.5 \pm 11.1\%$ of voxelwise response to breath holding; $t(15) = -0.54, p = .60$). The ventral bank response to Backwards was less than the dorsal bank response to Backwards ($27.0 \pm 17.2\%$; $t(15) = 5.34$; $p < .001$), and less than the ventral bank response to Spoken ($21.1 \pm 13.8\%$; $t(15) = 7.10$; $p < .001$).

In the right hemisphere, responses to the Spoken and Backwards conditions were much less extensive in the posterior STS and angular gyrus. As in the left hemisphere, backwards speech activated the dorsal bank of the STS. In the anterior STG, STS and MTG, a dorsal-to-ventral gradient was observed that was similar to that seen in the left hemisphere.

Written language processing

Brain regions involved in any aspect of written language comprehension were identified by plotting the union of two contrasts: Written – Rest (Fig. 2C, transparent), and Written – Scrambled (Fig. 2C, opaque). Again consistent with much prior research, the Written – Rest contrast activated bilateral occipital and posterior ventral temporal regions, and bilateral frontal and parietal regions, and the Written – Scrambled contrast activated bilateral regions in the anterior STS and MTG, and left-lateralized regions in the posterior STS, angular gyrus, and IFG.

We then plotted the response to the Scrambled – Rest contrast in the broad set of regions just identified. Lateral projections provide an overview in Fig. 2D, and a series of coronal slices through the left STS shows the details of patterns in Fig. 2E.

Just as in the auditory modality, a clear dorsal-to-ventral processing gradient was again apparent. All of the bilateral occipital, posterior ventral temporal, frontal and parietal regions that were identified in the Written – Rest contrast responded to scrambled text (Fig. 2D and E, uniform orange and red in transparent regions). These regions are involved in visual processing of the text and saccades for reading, as evidenced by their activation for Written – Rest but not Written – Scrambled. Moreover, they were activated for the contrast of Written – Spoken (Fig. 2D and E, gray outlines).

The dorsal bank of the STS responded to scrambled text. It was pointed out earlier that dorsal bank responses to backwards speech do not appear to reflect auditory processing; this is even more apparent with regard to scrambled text, which does not even involve the auditory modality. Dorsal bank responses to scrambled text may reflect processing of graphemic representations, orthographic word forms, and/or spreading activation to abstract lexical nodes, since these are the levels of written word processing that would be invoked by scrambled text, according to models of written word recognition. Most, but not all, dorsal

bank regions, responded even more strongly to intelligible text (i.e. they were also activated by the Written – Scrambled contrast).

Just as was the case for backwards speech, the ventral bank of the STS generally showed little to no response to scrambled text (green). Also similarly, scrambled text deactivated some ventral bank regions, the anterior MTG, and the angular gyrus relative to rest (blue). Most of the IFG did not respond to scrambled text (with the exception of the most dorsal parts of each of the two activations). The lack of response to scrambled text suggests that all of these regions are involved in higher level linguistic processing, and these regions were generally activated by the Written – Scrambled contrast.

To confirm the functional distinction between the dorsal and ventral banks of the STS, we examined the mean signal by condition in the dorsal and ventral bank anatomical ROIs. The ventral bank ROI showed a negative response to the Scrambled condition (mean signal change $= -9.7 \pm 14.1\%$ of voxelwise response to breath holding; $t(15) = -2.75, p = .015$). The ventral bank response to Scrambled was less than the dorsal bank response to Scrambled ($8.6 \pm 15.6\%$; $t(15) = 2.70$; $p = .016$), and less than the ventral bank response to Written ($35.4 \pm 13.9\%$; $t(15) = 5.77$; $p < .001$).

In the right hemisphere, responses to written and scrambled text were much less extensive in the posterior STS and angular gyrus, but scrambled speech did activate the dorsal bank of the STS. In the anterior STG, STS and MTG, a dorsal-to-ventral gradient was observed that was similar to that seen in the left hemisphere.

Convergence of spoken and written language processing

The gradients of processing that were observed in the STS, MTG, angular gyrus, and IFG were strikingly similar for spoken and written language processing. Three analyses were carried out to quantify these apparent similarities.

First, the regions involved in higher level processing of spoken and written language (i.e. the contrasts Spoken – Backwards, and Written – Scrambled) were directly compared. That is, the contrast (Spoken – Backwards) – (Written – Scrambled), and its inverse, were computed. There were no regions that were significantly activated by either of these contrasts, confirming that there were no differences between the regions involved in higher level processing of spoken and written language that could be not be excluded as being due to chance.

Second, to quantify the apparent similarity of the gradients of responses to the Backwards and Scrambled conditions in the STS and other brain regions (Figs. 1D, E and 2D, E), the contrasts Backwards – Scrambled, and Scrambled – Backwards, were computed. The regions activated by these contrasts were very similar to those already shown for Spoken – Written and its inverse (Figs. 1C–E and 2C,D, gray outlines). Regions that responded differentially to backwards speech or scrambled text were localized to early modality-specific regions, and showed almost no overlap with regions implicated in higher level processing in either modality. Specifically, 305 out of 15,900 (1.9%) voxels that were activated for higher level processing in either modality ((Spoken – Backwards) or (Written – Scrambled)) were also activated by Backwards – Scrambled or its inverse. These 305 voxels were located mostly in the dorsal bank of the anterior STS bilaterally, immediately adjacent to auditory areas on the lateral surface of the STG, and all showed more activity for backwards speech than scrambled text. Aside from these auditory-adjacent voxels, there were no other differences between the gradients shown in Figs. 1D, E and 2D,E that could not be excluded as being due to chance.

Third, given that mass univariate analyses failed to reveal any distinctions between spoken and written language processing except in or immediately adjacent to early modality-specific regions, searchlight MVPA was used to identify any regions within the broad language processing regions defined above (not including modality-specific regions) that could discriminate between the Spoken and Written conditions by means of distributed patterns of activity. This analysis showed that the

posterior STS was capable of discriminating between the Spoken and Written conditions based on multi-voxel patterns of activity (Fig. 3). Clusters were statistically significant in the left posterior STS (center of mass = −57, −48, 8, extent = 1255 mm^3, $p < .001$) and the right posterior STS (center of mass = 53, −41, 11; extent = 338 mm^3, $p = .008$). Across the left posterior STS cluster, the mean accuracy of multivariate classifiers (76.0 ± 12.7%) exceeded the accuracy of univariate mean signal classifiers (59.9 ± 8.7%) by 16.1 ± 7.4%. This ability to discriminate between the Spoken and Written conditions implies that the processing performed in this region is still modality-specific to some extent. In contrast, none of the other regions involved in language processing (subsequent to early modality-specific regions) were capable of discriminating between spoken and written language. While it cannot be excluded that studies using other classification approaches or with greater power might reveal sensitivity to input modality in other regions, the lack of such sensitivity in our study suggests that these other language regions may operate on amodal representations.

Discussion

The overall goal of this study was to better understand the functional and anatomical details of how spoken and written language processing streams converge in the STS. Our study had two paramount empirical findings. First, we observed that for both spoken and written language, there was a dorsal-to-ventral processing gradient within the STS, especially in the left hemisphere, such that there was a robust response to unintelligible stimuli in the dorsal bank, which was attenuated in the ventral bank. The striking similarity of this gradient across the auditory and visual modalities might suggest at first glance that all of the processing in the STS takes place subsequent to the convergence of the spoken and written language processing streams. However, this inference must be tempered by the second main empirical finding, which was that the posterior STS, unlike any other region implicated in higher level language processing, was capable of distinguishing between spoken and written inputs, based on patterns of activity distributed over multiple voxels, even though its net response to both input modalities was equivalent. This suggests that processing is actually not entirely amodal at that point.

Building on these observations, in the following sections we present our interpretation of the functional and anatomical specifics of how spoken and written language processing converge, which depends on a functional parcellation of the STS into a posterior dorsal bank region, an anterior dorsal bank region, and the regions activated on the ventral bank

and adjacent gyri. A schematic overview of our interpretation of our findings is presented in Fig. 4.

The posterior dorsal bank of the STS

The dorsal bank of the STS was the only brain region to be activated by intelligible as well as unintelligible inputs in both modalities, especially in the left hemisphere. However, we can go further in identifying the site of first convergence, because only the posterior part of the STS was also able to discriminate between spoken and written inputs based on distributed patterns of activity, indicating that this region is not yet modality-independent. In contrast, the anterior STS was not sensitive to the distinction between spoken and written inputs, suggesting that it is modality-independent. This positions the posterior dorsal bank as the first site at which spoken and written inputs converge.

There is a clear anatomical basis for convergence in the posterior dorsal bank of the STS. In non-human primates, auditory inputs to the STS are largely restricted to the dorsal bank (Seltzer and Pandya, 1978; Cusick et al., 1995; Seltzer et al., 1996; Hackett et al., 2014). The main site of multisensory convergence is the caudal part of area TPO, located on the dorsal bank of the STS (Seltzer and Pandya, 1978; Desimone and Ungerleider, 1986; Cusick et al., 1995; Padberg et al., 2003). While homologies between monkeys and humans have not been clearly established, these findings are highly consistent with the present evidence for convergence of spoken and written language processing in the posterior dorsal bank. In humans, diffusion tensor imaging studies have demonstrated connectivity between auditory regions and the STS (Beer et al., 2013), and between visual regions and the STS (Beer et al., 2013), including regions specifically involved in reading (Bouhali et al., 2014). To our knowledge, no studies have attempted to resolve auditory or visual projections to the dorsal or ventral bank specifically.

Previous functional imaging studies have established that spoken and written language processing converge in the STS (van Atteveldt et al., 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Berl et al., 2010; Vagharchakian et al., 2012; Regev et al., 2013). However, these studies did not have the spatial resolution to resolve the dorsal and ventral banks of the STS. Moreover, all but one of these studies (Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Berl et al., 2010; Vagharchakian et al., 2012; Regev et al., 2013) were essentially dependent on contrasts of intelligibility, which means that common activations between spoken and written language processing would include regions involved in lexical, semantic and syntactic processing, which are downstream of the site of first convergence.
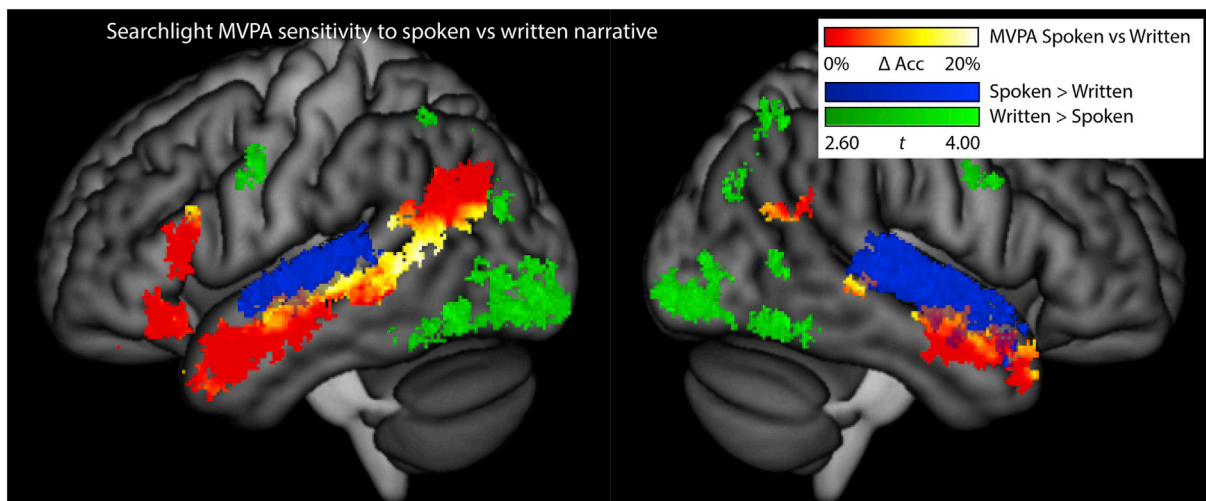

Fig. 3. Multi-voxel discrimination between spoken and written language. The hot color scale shows the increase in accuracy for a multivariate searchlight classifier compared to a univariate classifier, in regions involved in language processing, except for regions that were demonstrated to be modality-specific in univariate analyses (these are shown in blue for auditory areas, and green for visual areas).
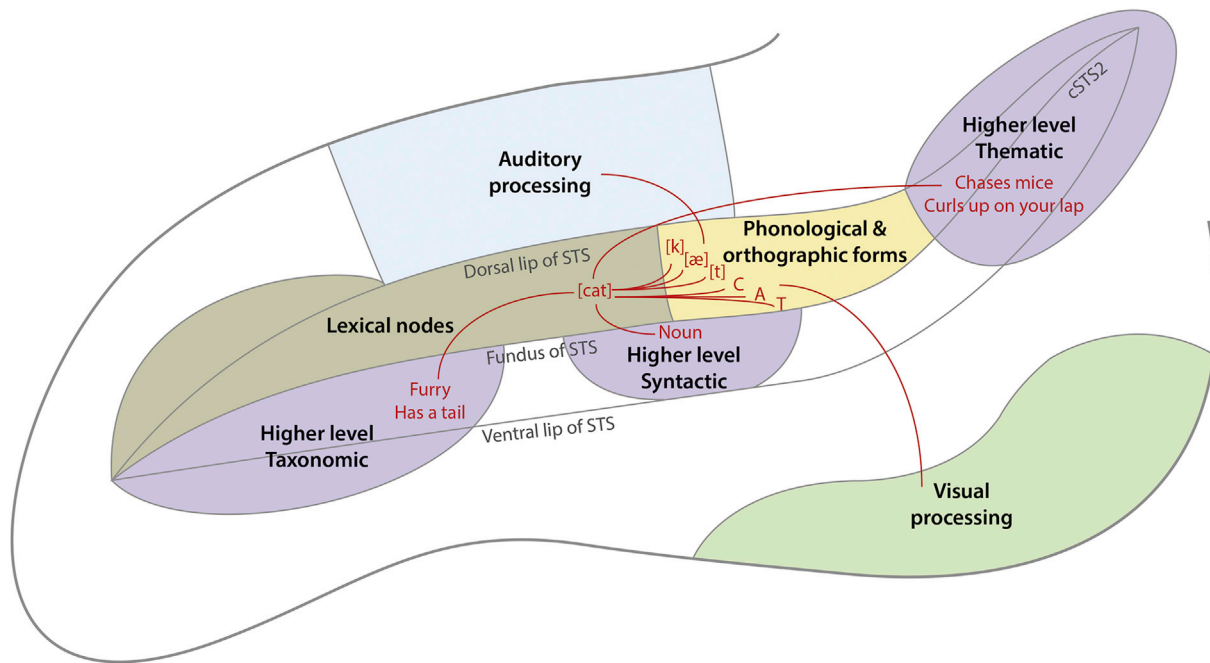
Fig. 4. Schematic overview of the heterogeneous language regions within the STS. The key elements of our model are: (1) the processing of spoken and written language converges on the posterior dorsal bank of the STS (yellow); (2) this region is involved in representing phonological and orthographic word forms; (3) abstract lexical nodes are encoded in the anterior dorsal bank of the STS (olive); (4) higher level semantic and syntactic processing depends on three regions located on the ventral bank of the STS and adjacent gyri (purple). An example lexical item ('cat') is shown in red.

Accordingly, these studies all reported sites of convergence along the whole length of the STS. The study by van Atteveldt et al. (2004) was the exception, since it involved only low-level stimuli (single letters and phonemes). The most prominent region of convergence in that study was the posterior STS, consistent with our findings. Also noteworthy is another important type of auditory-visual convergence that is relevant to language comprehension: the role of visual information (the talker's mouth) in speech perception. Visual information changes the auditory percept at a pre-lexical level (McGurk and MacDonald, 1976), and this type of integration of auditory and visual information has also been shown to take place in the posterior STS (Calvert et al., 2000; Nath and Beauchamp, 2012).

What is the nature of the representations encoded in the posterior dorsal bank region? Models of spoken and written word recognition generally posit layers of phonological and graphemic representations respectively, which are linked to lexical representations via a spreading activation mechanism (McClelland and Rumelhart, 1981; McClelland and Elman, 1986). According to these models, the unintelligible conditions in our study (Backwards and Scrambled) should activate phonological and orthographic representations by virtue of the partial phonemic information and the letters of which they are comprised. Moreover, these inputs would also be expected to activate lexical nodes via spreading activation, even though no lexical node would ultimately be selected. Simulation evidence in the orthographic modality is consistent with this assumption (Taylor et al., 2013).

Because the posterior dorsal bank responded to both spoken and written inputs, it seems unlikely to represent lower level constructs such as phonological features, phonemes or graphemes that are unique to each modality. Moreover, there is evidence that auditory information is already at least partially shaped into phonemic categories earlier, on the lateral surface of the STG (Chang et al., 2010), and that abstract letter representations are encoded in the visual word form area (VWFA) at the anterior extent of the occipito-temporal visual word form system (Dehaene and Cohen, 2011). On the other hand, the posterior dorsal bank was capable of differentiating between spoken and written inputs, so it seems unlikely to encode modality-neutral representations, such as

amodal lexical nodes, and the fact that it responded to unintelligible inputs suggests that it does not encode semantic or syntactic representations. Situated in between these levels are representations of phonological and orthographic word forms. This seems to be the most plausible type of information that is encoded on the posterior dorsal bank of the STS.

Specifically, we propose that phonological word forms are represented through patterns of connections between phonemic representations in higher level auditory regions on the lateral surface of the STG, and distributed patterns of activity on the posterior dorsal bank of the STS. The view that phonological word forms are localized to the mid or posterior STS is widely held (Indefrey and Levelt, 2004; Okada and Hickok, 2006; Hickok and Poeppel, 2007; Vaden et al., 2010), but to our knowledge no prior studies have localized these representations to the dorsal bank specifically.

Similarly, we propose that orthographic word forms are represented via patterns of connections between abstract graphemic representations in the VWFA, and distributed patterns of activity in the same STS region. An alternative explanation is that posterior dorsal bank activation for written language comprehension reflects covert phonological mediation or inner speech (Leinenger, 2014), however we consider this less likely, because the dorsal bank was activated even in the scrambled condition, in which most of the scrambled words were not even pronounceable (Fig. 2B).

The localization of orthographic word forms is controversial. Some researchers have claimed that orthographic word forms have distinct cortical substrates from phonological word forms (Howard et al., 1992; Booth et al., 2002; Taylor et al., 2013), while others have argued that there is no evidence for distinct cortical sites (Price et al., 2003). Of the various regions proposed to be involved in representation of orthographic word forms, the most compelling candidate is probably the VWFA, at the anterior culmination of the occipito-temporal visual word form system. There are several pieces of evidence that suggest the VWFA could play a role in storage of orthographic word forms. First, lesions to the VWFA have been shown to result not only in reading deficits, but also spelling deficits, especially for irregular words, which are especially

reliant on access to orthographic word forms (Rapcsak and Beeson, 2004). Second, some imaging studies have reported differential activation in this region for processing orthographically inconsistent words, which again would make explicit demands on stored orthographic forms (Graves et al., 2010). Third, in a meta-analysis, the anterior part of the VWFA was shown to be activated for words, which have orthographic word forms, relative to nonwords, which do not (Taylor et al., 2013).

In our study, the VWFA was not activated by the contrast between written and scrambled text. This constitutes evidence, but not strong evidence, against a role for this region in representation of orthographic word forms. The reason that this finding is far from conclusive is that responses to text and scrambled text in the VWFA (and responses to degraded stimuli in the brain in general) reflect a complex interplay of degree of engagement and degree of processing effort, both of which are strongly influenced by specific task demands (Taylor et al., 2013, 2014). In some experimental contexts, such as rapid presentation of unrelated words, the VWFA certainly is sensitive to the distinction between real words and meaningless letter strings (Vinckier et al., 2007). But even in that study, the VWFA did not distinguish between real words and phonotactically well-formed nonwords. In our view, the weight of the evidence suggests that the VWFA is responsible for *pre-lexical* representation of the visual word form input (Dehaene and Cohen, 2011). The situations cited above in which the VWFA has been associated with lexical factors may reflect the direct links between pre-lexical representations in the VWFA and stored orthographic word forms in the STS, which may be a source of top down feedback (Price and Devlin, 2011).

There has been much debate as to whether linguistic processing of auditory inputs is initially directed anteriorly (Scott et al., 2000; Obleser et al., 2006; Rauschecker and Scott, 2009; DeWitt and Rauschecker, 2012; Evans et al., 2014) or posteriorly (Kertesz et al., 1982; Selnes et al., 1983; Hickok and Poeppel, 2007; Okada et al., 2010). Our finding that the posterior STS is sensitive to input modality, while the anterior STS is amodal, favors the view that processing is initially directed posteriorly. While there is compelling evidence for an anteriorly oriented auditory object identification pathway in non-human primates (Rauschecker and Scott, 2009), it is not clear that language is an "auditory object" in a sense that would dictate dependence on the same neural substrates. In contrast, identification of speaker identity based on voice seems more akin to "auditory object" identification, and is more clearly dependent on an anterior temporal pathway (Belin and Zatorre, 2003).

The anterior dorsal bank of the STS

The anterior part of the dorsal bank of the STS responded to unintelligible inputs in both modalities. Yet unlike the posterior STS, distributed patterns of activity in this region did not discriminate between modalities, suggesting that it is modality-independent. What is the nature of the representations encoded in this region? Its insensitivity to input modality argues against representations tied to one modality or the other, such as phonemes, graphemes, or phonological or orthographic word forms. On the other hand, the fact that it responded to unintelligible inputs suggests that it does not encode semantic or syntactic representations. This narrows down the most likely type of information encoded on the anterior dorsal bank of the STS to be amodal lexical nodes, akin to the concept of 'lemma' that is more prominent in speech production research.

We propose that the representations in this region are distributed patterns of activity that encode abstract lexical entries, but do not contain phonological, orthographic, semantic or syntactic information. Rather, these lexical nodes serve as hubs that bind together these different types of information, which are predominantly stored and represented in other brain regions: phonological and orthographic word forms in the posterior dorsal bank of the STS as just proposed, and semantic and syntactic information in ventral bank and adjacent regions to be discussed below. Theories of the neural basis of language comprehension have not tended to make any reference to lemma-like concepts, instead modeling

direct links between phonological and semantic representations (e.g. Hickok and Poeppel, 2007; Binder, 2015). However, our anterior temporal localization of lemma-like representations in comprehension is concordant with studies of lexical retrieval and speech production (de Zubicaray et al., 2001; Indefrey and Levelt, 2004; Schwartz et al., 2009). The consequences of damage to this region are informative. Schwartz et al. (2009) showed that damage to the left anterior temporal region is associated with semantic errors (e.g. misnaming 'elephant' as 'zebra') above and beyond impairment of semantic knowledge. In other words, damage interfered with the links between meanings and word forms, not with meanings or word forms themselves. Along the same lines, Lambon Ralph et al. (2001) and Wilson et al. (2017) have shown that semantic dementia patients with left-lateralized anterior temporal atrophy have naming deficits that cannot be accounted for solely in terms of underlying semantic impairment, and must also involve damage to links between semantic representations and lemmas or left-lateralized word form representations. Note also that phonemic errors are very rare in semantic dementia (Hodges et al., 1992; Wilson et al., 2010b), consistent with a more posterior locus for representation of word forms themselves.

The ventral bank of the STS, and adjacent regions

Several regions on the ventral bank of the STS, and adjacent regions, were activated equivalently by spoken and written language, but not by the unintelligible conditions: backwards speech or scrambled text. This response pattern is suggestive of higher level linguistic processing, including encoding of semantic and syntactic representations. Higher level stages of processing such as these would not be invoked by unintelligible stimuli, which do not result in the selection of any lexical node, and hence do not activate representations of meaning (Connine et al., 1993; Perea and Lupker, 2003). Indeed, some parts of these regions actually showed less activity for unintelligible inputs than at rest; this can be interpreted as attenuation by meaningless stimuli of semantic processing that takes place during the resting state (Binder et al., 1999).

Closer examination of these modality-independent regions that responded only to intelligible stimuli reveals three distinct clusters, which can be observed independently in the spoken and written contrasts. First, there was an anterior region on the ventral bank of the STS and adjacent MTG, which can be observed in slices $y = 14$ through $y = -18$ in Figs. 1 and 2; this area was largely bilateral. Second, there was a mid-posterior region on the ventral bank of the STS, that can be observed on slices $y = -30$ through $y = -42$ in Figs. 1 and 2; this region was markedly left-lateralized. Third, there was a posterior region in the middle branch of the caudal STS (cSTS2) and the adjacent angular gyrus; this can be observed on slices $y = -54$ through $y = -66$ in Figs. 1 and 2. This region actually extended to both banks of this sulcal branch, but was still anatomically "downstream" (medial) relative to the lower level activations. Like the mid-posterior STS region, this region was markedly left-lateralized. Note that the posterior STS region discussed above that could discriminate between spoken and written language based on multivoxel patterns lay between these second and third higher level regions, in a part of the STS where there was no ventral bank response to any contrast. Thus no higher level regions showed any ability to differentiate between spoken and written inputs.

Although the present study provides no basis for distinguishing functionally between the three higher level regions observed, a consideration of other relevant literature suggests that they have distinct functions. The anterior temporal and angular gyrus regions may be associated with "taxonomic" and "thematic" aspects of semantic representations respectively (Hodges et al., 1992; Wu et al., 2007; Schwartz et al., 2011; Thothathiri et al., 2012), while the mid-posterior STS region appears to be particularly important for syntactic processing (Friederici et al., 2009; Pallier et al., 2011; Wilson et al., 2016).

The higher level areas we identified are quite similar to a left-lateralized semantic network that has been demonstrated in many studies, as shown in a critical review and meta-analysis (Binder et al.,

2009). Two anatomical details deserve comment. First, while previous studies have generally described the MTG and angular gyrus as comprising nodes of the semantic network, we were able to show that these regions are centered on the ventral bank of the STS, and both banks of cSTS2, from where they extend onto the gyri in question. Second, we did not observe activations related to higher level processing to extend onto the lateral surface of the posterior MTG, in contrast to Binder et al. (2009). We speculate that the posterior MTG is indeed a critical language region but that it is involved in language production and controlled tasks, and less so in language comprehension. This conjecture is based on the functional connectivity of this region: it forms a strongly left-lateralized network with inferior frontal and inferior parietal regions (Smith et al., 2009; Turken and Dronkers, 2011) that are modulated in parallel in language production (Geranmayeh et al., 2012).

Limitations and future directions

Our study had several noteworthy limitations, which could be addressed in future work. First, hypercapnic normalization and masking out of veins provide only a partial solution to the challenges of fine localization that stem from vascular anatomy and the physiology of the BOLD effect (Bandettini and Wong, 1997; Cohen et al., 2004; Handwerker et al., 2007; Thomason et al., 2007; Murphy et al., 2011; Wilson, 2014). Techniques such as arterial spin labeling or spin-echo BOLD are more specific to parenchymal signal, though at the cost of reduced sensitivity.

Second, we focused on group results rather than looking at individual participants. Therefore, the patterns we found may reflect interindividual variation as well as processing gradients in individual participants. In a previous study, we showed that it is feasible to localize activations to the dorsal or ventral banks of the STS in individual participants (Wilson, 2014), but the relatively lower power in each individual participant resulted in ventral bank activity being underestimated in that study. Multi-session studies of individual participants could alleviate this problem and yield highly accurate maps.

Third, we employed only one type of unintelligible stimuli in each modality. It could be informative to investigate responses to a range of inputs varying in the extent to which they would be expected to invoke different stages of processing, similar to Vinckier et al.'s (2007) study of single word reading, which involved false fonts, four types of nonwords varying in phonotactic similarity to real words, and real words. The rapid presentation mode and non-linguistic task employed in that study highlighted a gradient of processing in the visual word form system. A similar study based on narrative stimuli may shed light on the functional parcellation of the wider language network. Interpretation would be challenging however, because as mentioned above, neural responses to degraded stimuli reflect a complex interplay of engagement, processing effort, and task demands (Taylor et al., 2013, 2014).

Fourth, our study had only sixteen participants. While this is comparable to previous studies that have demonstrated linguistic processing gradients (Davis and Johnsrude, 2003; Vinckier et al., 2007; Visser et al., 2012), we cannot exclude that some null results may reflect lack of power. An *a priori* power analysis was precluded because the main patterns of interest in our study (that is, similar processing gradients for spoken and written language, and differences between regions in their ability to discriminate between spoken and written stimuli) are not of a form that would be amenable to power analysis, nor were estimated effect sizes available (Mumford and Nichols, 2008). In view of this limitation, it is highly plausible that there are subtle differences between regions involved in higher level processing of spoken and written language such that the contrast (Spoken – Backwards) – (Written – Scrambled) and its inverse would yield activation(s) with a larger group of participants. Similarly, there may be language regions other than the posterior STS that are capable of discriminating between spoken and written language.

Conclusion

By investigating neural responses to intelligible and unintelligible inputs in spoken and written modalities, using univariate and multivariate approaches, and implementing a range of strategies to maximize spatial resolution, we clarified the functional neuroanatomy of how spoken and written language processing converge in the STS. We found that the processing of spoken and written language converges on the posterior dorsal bank of the posterior STS, a region that responded to intelligible as well as unintelligible spoken and written inputs, and was sensitive to the distinction between spoken and written inputs based on distributed patterns of activity. Based on this functional profile, we argued that this region encodes phonological and orthographic word forms. The anterior dorsal bank of the STS also responded to intelligible and unintelligible inputs in both modalities, yet it was modality-neutral, showing no sensitivity to the distinction between spoken and written inputs. We argued that this suggests a role in representation of amodal lemma-like nodes that mediate between word forms and higher level representations. Several regions on the ventral bank of the STS and adjacent gyri responded only to intelligible inputs and were modality-neutral, consistent with involvement in higher level semantic and/or syntactic processes. Taken together, our findings show that there are a heterogeneous set of language regions within the STS, with distinct functions spanning a broad spectrum of linguistic processes and representations.

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bandettini, P.A., Wong, E.C., 1997. A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. NMR Biomed. 10, 197–203.

Bauer, J., 2004. Hope Was Here [compact disc]. Lamua, J., reader. Random House/Listening Library, New York.

Beer, A.L., Plank, T., Meyer, G., Greenlee, M.W., 2013. Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. Front. Integr. Neurosci. 7, 5.

Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105–2109.

Berl, M.M., Duke, E.S., Mayo, J., Rosenberger, L.R., Moore, E.N., VanMeter, J., Ratner, N.B., Vaidya, C.J., Gaillard, W.D., 2010. Functional anatomy of listening and reading comprehension during development. Brain Lang. 114, 115–125.

Binder, J., 2000. The new neuroanatomy of speech perception. Brain 123, 2371–2372.

Binder, J.R., 2015. The Wernicke area: modern evidence and a reinterpretation. Neurology 85, 2170–2175.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebr. Cortex 19, 2767–2796.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Rao, S.M., Cox, R.W., 1999. Conceptual processing during the conscious resting state: a functional MRI study. J. Cogn. Neurosci. 11, 80–93.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cerebr. Cortex 10, 512–528.

Binder, J.R., Frost, J.A., Hammeke, T.A., Rao, S.M., Cox, R.W., 1996. Function of the left planum temporale in auditory and linguistic processing. Brain 119, 1239–1247.

Binder, J.R., Mohr, J.P., 1992. The topography of callosal reading pathways: a case-control analysis. Brain 115, 1807–1826.

Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2002. Functional anatomy of intra- and cross-modal lexical tasks. NeuroImage 16, 7–22.

Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J.-F., Dehaene, S., Cohen, L., 2014. Anatomical connections of the visual word form area. J. Neurosci. 34, 15402–15414.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432.

Choi, W., Desai, R.H., Henderson, J.M., 2014. The neural substrates of natural reading: a comparison of normal and nonword text using eyetracking and fMRI. Front. Hum. Neurosci. 8, 1024.

Cohen, E.R., Rostrup, E., Sidaros, K., Lund, T.E., Paulson, O.B., Ugurbil, K., Kim, S.-G., 2004. Hypercapnic normalization of BOLD fMRI: comparison across field strengths and pulse sequences. NeuroImage 23, 613–624.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123, 291–307.

Connine, C.M., Blasko, D.G., Titone, D., 1993. Do the beginnings of spoken words have a special status in auditory word recognition? J. Mem. Lang. 32, 193–210.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Cusick, C.G., Seltzer, B., Cola, M., Griggs, E., 1995. Chemoarchitectonics and corticocortical terminations within the superior temporal sulcus of the rhesus monkey: evidence for subdivisions of superior temporal polysensory cortex. J. Comp. Neurol. 360, 513–535.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.

de Zubicaray, G.I., Wilson, S.J., McMahon, K.L., Muthiah, S., 2001. The semantic interference effect in the picture-word paradigm: an event-related fMRI study employing overt responses. Hum. Brain Mapp. 14, 218–227.

Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn Sci. 15, 254–262.

Desimone, R., Ungerleider, L.G., 1986. Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol. 248, 164–189.

DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. 109, E505–E514.

Evans, S., Kyong, J.S., Rosen, S., Golestani, N., Warren, J.E., McGettigan, C., Mourao-Miranda, J., Wise, R.J.S., Scott, S.K., 2014. The pathways for intelligible speech: multivariate and univariate perspectives. Cerebr. Cortex 24, 2350–2361.

Formisano, E., Kim, D.S., Di Salle, F., van de Moortele, P.F., Ugurbil, K., Goebel, R., 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40, 859–869.

Friederici, A.D., Makuuchi, M., Bahlmann, J., 2009. The role of the posterior superior temporal cortex in sentence comprehension. Neuroreport 20, 563–568.

Friston, K.J., 2007. Statistical Parametric Mapping the Analysis of Funtional Brain Images. Elsevier/Academic Press, Amsterdam.

Geranmayeh, F., Brownsett, S.L.E., Leech, R., Beckmann, C.F., Woodhead, Z., Wise, R.J.S., 2012. The contribution of the inferior parietal cortex to spoken language production. Brain Lang. 121, 47–57.

Geschwind, N., 1965. Disconnexion syndromes in animals and man. I. Brain 88, 237–294.

Graves, W.W., Desai, R., Humphries, C., Seidenberg, M.S., Binder, J.R., 2010. Neural systems for reading aloud: a multiparametric approach. Cerebr. Cortex 20, 1799–1815.

Hackett, T.A., de la Mothe, L.A., Camalier, C.R., Falchier, A., Lakatos, P., Kajikawa, Y., Schroeder, C.E., 2014. Feedforward and feedback projections of caudal belt and parabelt areas of auditory cortex: refining the hierarchical model. Front. Neurosci. 8, 72.

Handwerker, D.A., Gazzaley, A., Inglis, B.A., D'Esposito, M., 2007. Reducing vascular variability of fMRI data across aging populations using a breathholding task. Hum. Brain Mapp. 28, 846–859.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Hillen, R., Günther, T., Kohlen, C., Eckers, C., van Ermingen-Marbach, M., Sass, K., Scharke, W., Vollmar, J., Radach, R., Heim, S., 2013. Identifying brain systems for gaze orienting during reading: fMRI investigation of the Landolt paradigm. Front. Hum. Neurosci. 7, 384.

Hodges, J.R., Patterson, K., Oxbury, S., Funnell, E., 1992. Semantic dementia: progressive fluent aphasia with temporal lobe atrophy. Brain 115, 1783–1806.

Howard, D., Patterson, K., Wise, R., Brown, W.D., Friston, K., Weiller, C., Frackowiak, R., 1992. The cortical localization of the lexicons: positron emission tomography evidence. Brain 115, 1769–1782.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458.

Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144.

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.

Jones, E.G., Powell, T.P., 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93, 793–820.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kertesz, A., Sheppard, A., MacKenzie, R., 1982. Localization in transcortical sensory aphasia. Arch. Neurol. 39, 475–478.

Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V., 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868.

Lai, S., Hopkins, A.L., Haacke, E.M., Li, D., Wasserman, B.A., Buckley, P., Friedman, L., Meltzer, H., Hedera, P., Friedland, R., 1993. Identification of vascular structures as a major source of signal contrast in high resolution 2D and 3D functional activation imaging of the motor cortex at 1.5T: preliminary results. Magn. Reson. Med. 30, 387–392.

Lambon Ralph, M.A., McClelland, J.L., Patterson, K., Galton, C.J., Hodges, J.R., 2001. No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. J. Cogn. Neurosci. 13, 341–356.

Leinenger, M., 2014. Phonological coding during reading. Psychol. Bull. 140, 1534–1555.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cerebr. Cortex 15, 1621–1631.

Liebenthal, E., Desai, R.H., Humphries, C., Sabri, M., Desai, A., 2014. The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. Front. Neurosci. 8, 289.

Lindenberg, R., Scheef, L., 2007. Supramodal language comprehension: role of the left temporal lobe for listening and reading. Neuropsychologia 45, 2407–2415.

McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn Psychol 18, 1–86.

McClelland, J.L., Rumelhart, D.E., 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. Psych. Rev. 88, 375–407.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

Menon, R.S., Ogawa, S., Tank, D.W., Uğurbil, K., 1993. 4 Tesla gradient recalled echo characteristics of photic stimulation-induced signal changes in the human primary visual cortex. Magn. Reson. Med. 30, 380–386.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343, 1006–1010.

Moerel, M., De Martino, F., Formisano, E., 2014. An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. 8, 225.

Möttönen, R., Calvert, G.A., Jääskeläinen, I.P., Matthews, P.M., Thesen, T., Tuomainen, J., Sams, M., 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. NeuroImage 30, 563–569.

Mumford, J.A., Nichols, T.E., 2008. Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. NeuroImage 39, 261–268.

Murphy, K., Harris, A.D., Wise, R.G., 2011. Robustly measuring vascular reactivity differences with breath-hold: normalising stimulus-evoked and resting state BOLD fMRI data. NeuroImage 54, 369–379.

Nath, A.R., Beauchamp, M.S., 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage 59, 781–787.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Obleser, J., Boecker, H., Drzezga, A., Haslinger, B., Hennenlotter, A., Roettinger, M., Eulitz, C., Rauschecker, J.P., 2006. Vowel sound extraction in anterior superior temporal cortex. Hum. Brain Mapp. 27, 562–571.

Okada, K., Hickok, G., 2006. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. Neuroreport 17, 1293–1296.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cerebr. Cortex 20, 2486–2495.

Padberg, J., Seltzer, B., Cusick, C.G., 2003. Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. J. Comp. Neurol. 467, 418–434.

Pallier, C., Devauchelle, A.D., Dehaene, S., 2011. Cortical representation of the constituent structure of sentences. Proc. Natl. Acad. Sci. U. S. A. 108, 2522–2527.

Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–987.

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.

Perea, M., Lupker, S.J., 2003. Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. Mem. Cognit. 31, 829–841.

Pereira, F., Botvinick, M., 2011. Information mapping with pattern classifiers: a comparative study. NeuroImage 56, 476–496.

Price, C.J., Devlin, J.T., 2011. The interactive account of ventral occipitotemporal contributions to reading. Trends Cogn. Sci. 15, 246–253.

Price, C.J., Winterburn, D., Giraud, A.L., Moore, C.J., Noppeney, U., 2003. Cortical localisation of the visual and auditory word form areas: a reconsideration of the evidence. Brain Lang. 86, 272–286.

Rapcsak, S.Z., Beeson, P.M., 2004. The role of left posterior inferior temporal cortex in spelling. Neurology 62, 2221–2229.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724.

Regev, M., Honey, C.J., Simony, E., Hasson, U., 2013. Selective and invariant neural responses to spoken and written narratives. J. Neurosci. 33, 15978–15988.

Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4 (1), 32.

Schwartz, M.F., Kimberg, D.Y., Walker, G.M., Brecher, A., Faseyitan, O.K., Dell, G.S., Mirman, D., Coslett, H.B., 2011. Neuroanatomical dissociation for taxonomic and thematic knowledge in the human brain. Proc. Natl. Acad. Sci. U. S. A. 108, 8520–8524.

Schwartz, M.F., Kimberg, D.Y., Walker, G.M., Faseyitan, O., Brecher, A., Dell, G.S., Coslett, H., 2009. Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. Brain 132, 3411–3427.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Segal, E., Petrides, M., 2012. The morphology and variability of the caudal rami of the superior temporal sulcus. Eur. J. Neurosci. 36, 2035–2053.

Selnes, O.A., Knopman, D.S., Niccum, N., Rubens, A.B., Larson, D., 1983. Computed tomographic scan correlates of auditory comprehension deficits in aphasia: a prospective recovery study. Ann. Neurol. 13, 558–566.

Seltzer, B., Cola, M.G., Gutierrez, C., Massee, M., Weldon, C., Cusick, C.G., 1996. Overlapping and nonoverlapping cortical projections to cortex of the superior temporal sulcus in the rhesus monkey: double anterograde tracer studies. J. Comp. Neurol. 370, 173–190.

Seltzer, B., Pandya, D.N., 1978. Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res. 149, 1–24.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain's functional architecture during activation and rest. Proc. Natl. Acad. Sci. U. S. A. 106, 13040–13045.

Spitsyna, G., Warren, J.E., Scott, S.K., Turkheimer, F.E., Wise, R.J.S., 2006. Converging language streams in the human temporal lobe. J. Neurosci. 26, 7328–7336.

Stelzer, J., Buschmann, T., Lohmann, G., Margulies, D.S., Trampel, R., Turner, R., 2014. Prioritizing spatial accuracy in high-resolution fMRI data using multivariate feature weight mapping. Front. Neurosci. 8, 66.

Taylor, J.S.H., Rastle, K., Davis, M.H., 2013. Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. Psychol. Bull. 139, 766–791.

Taylor, J.S.H., Rastle, K., Davis, M.H., 2014. Interpreting response time effects in functional imaging studies. NeuroImage 99, 419–433.

Thomason, M.E., Foland, L.C., Glover, G.H., 2007. Calibration of BOLD fMRI using breath holding reduces group variance during a cognitive task. Hum. Brain Mapp. 28, 59–68.

Thothathiri, M., Kimberg, D.Y., Schwartz, M.F., 2012. The neural basis of reversible sentence comprehension: evidence from voxel-based lesion symptom mapping in aphasia. J. Cogn. Neurosci. 24, 212–222.

Turkeltaub, P.E., Coslett, H.B., 2010. Localization of sublexical speech perception components. Brain Lang. 114, 1–15.

Turken, A.U., Dronkers, N.F., 2011. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. Front. Syst. Neurosci. 5, 1.

Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W., Patterson, R.D., 2006. Locating the initial stages of speech-sound processing in human temporal cortex. NeuroImage 31, 1284–1296.

Vaden, K.I., Muftuler, L.T., Hickok, G., 2010. Phonological repetition-suppression in bilateral superior temporal sulci. NeuroImage 49, 1018–1023.

Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., Dehaene, S., 2012. A temporal bottleneck in the language comprehension network. J. Neurosci. 32, 9089–9102.

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., 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.

Visser, M., Jefferies, E., Embleton, K.V., Lambon Ralph, M.A., 2012. Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. J. Cogn. Neurosci. 24, 1766–1778.

Wilson, S.M., 2014. The impact of vascular factors on language localization in the superior temporal sulcus. Hum. Brain Mapp. 35, 4049–4063.

Wilson, S.M., Dehollain, C., Ferrieux, S., Christensen, L.E.H., Teichmann, M., 2017. Lexical access in semantic variant PPA: evidence for a post-semantic contribution to naming deficits. Neuropsychologia 106, 90–99.

Wilson, S.M., DeMarco, A.T., Henry, M.L., Gesierich, B., Babiak, M., Miller, B.L., Gorno-Tempini, M.L., 2016. Variable disruption of a syntactic processing network in primary progressive aphasia. Brain 139, 2994–3006.

Wilson, S.M., Dronkers, N.F., Ogar, J.M., Jang, J., Growdon, M.E., Agosta, F., Henry, M.L., Miller, B.L., Gorno-Tempini, M.L., 2010a. Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. J. Neurosci. 30, 16845–16854.

Wilson, S.M., Henry, M.L., Besbris, M., Ogar, J.M., Dronkers, N.F., Jarrold, W., Miller, B.L., Gorno-Tempini, M.L., 2010b. Connected speech production in three variants of primary progressive aphasia. Brain 133, 2069–2088.

Woodhead, Z.V.J., Brownsett, S.L.E., Dhanjal, N.S., Beckmann, C., Wise, R.J.S., 2011. The visual word form system in context. J. Neurosci. 31, 193–199.

Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.

Wu, D.H., Waller, S., Chatterjee, A., 2007. The functional neuroanatomy of thematic role and locative relational knowledge. J. Cogn. Neurosci. 19, 1542–1555.

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.