

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2017

Shared and Selective Neural Processes Across the Visual and Auditory Language Modalities

Regev, Mor

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01n583xx59k>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**Shared and Selective Neural Processes Across the Visual and
Auditory Language Modalities**

Mor Regev

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PSYCHOLOGY

Advisor: Uri Hasson

June 2017

© Copyright by Mor Regev, 2017. All rights reserved.

Abstract

Linguistic content can be conveyed both in speech and in writing. But how does the human brain accommodate the much later evolved skill of literacy—visual language—on top of the previously developed ability to comprehend auditory speech? This dissertation seeks to explore which neural processes are common and which diverge across the auditory and visual language modalities. Using functional magnetic resonance imaging, we pursued four goals: 1) map areas that reliably activate during reading of complex written narrative, and identify which of these regions are unique to reading and not to speaking; 2) map the regions that show functional similarities when reading and when listening to the same narrative; 3) examine the extent to which top-down attention modulates response to text and speech along their processing hierarchies; and 4) identify the neural pathways through which written and spoken attended and unattended content is distributed through the brain. First, we used inter-subject correlation to map the extended reading network that activates reliably during text comprehension. Early sensory areas and some high-order parietal and frontal areas responded selectively to written and spoken versions of the narrative. By contrast, the temporal response profiles in some linguistic and extra-linguistic areas were remarkably similar for spoken and written narratives, indicating strong modality-invariance of linguistic processing in these circuits. Next, we revealed a similar hierarchical modulation effect of attention on spoken and written content, in which sensory cortices processed unattended language, unlike high-order cognitive cortices. Further, the neural communication between brain regions was examined using inter-subject functional correlation analysis, which demonstrated that although story-related responses were shared between areas in sensory cortices even when outside the focus of attention, top-down attention was required for the narrative-specific response to be shared with higher, more cognitive frontal and parietal

areas. Overall, the results presented in this dissertation argue that our ability to extract the same information from spoken and written forms arises from a mixture of both selective and shared neural processes. This constitutes a major step in the understanding of the neurophysiological and functional dynamics underlying the coexistence of written and spoken language in the brain.

Acknowledgments

I would like to first express my sincere gratitude to my advisor, Uri Hasson, for his continuous support of my research, for the useful discussions and brainstorming sessions, and for his immense patience. His mentorship has been invaluable in shaping my scientific thinking. I would also like to take this opportunity to thank the members of my committee—Asif Ghazanfar, Ken Norman, Nick Turk-Browne, and Adele Goldberg—for their very helpful comments and suggestions. Janice Chen and Chris Honey have been a great source of knowledge and encouragement throughout all my years at Princeton, and I will always be grateful to them. A special thank you goes to Erez Simony for his expert advice, to Yaara Yeshurun-Dishon for her thoughtful insights, and to Chris Baldassano for his input. Thanks are also due to the entire Hasson lab, especially to Asieh Zadbood and Mai Nguyen, for stimulating exchanges of ideas, for their camaraderie and friendship. I thank Amir Amedi and Raz Yirmiya of the Hebrew University for taking a risk and opening their labs to me as an inexperienced undergraduate student, giving me my first valuable research experience. David Heeger, Eric Greenleaf, Vicki Morwitz, and Moran Cerf of New York University have likewise encouraged me during my first steps as a researcher.

I gratefully acknowledge the support and generosity of The Dingwall Foundation for their dissertation fellowship, without which this dissertation could not have been completed. I am thankful for the support of Princeton's Program in Cognitive Science for their graduate fellowship and research funding as well. I am also in debt to the entire Psychology and PNI staff members, who helped me navigate the intricate academic world. A special thanks goes to the MR staff who skillfully supported and enabled my research work. A big thank you also goes out to everybody who took part in my experiments, including all the participants and buddies, who

made the scanning sessions not just possible but also enjoyable. I thankfully acknowledge the hardworking crews of New Jersey Transit and the Dinky, who did their very best to enable me to divide my life between Princeton and the beautiful city of New York.

For their encouragement and fun times, I want to thank my brilliant Princeton friends. An exceptional thank you goes to Princeton's Graduate Housing Room Draw for giving me Reut Harari, and to Reut Harari for being my home in Princeton. Her friendship is one of the most valuable things I will carry with me from Princeton. Finally, my biggest gratitude goes to my beloved family. They have been a constant source of strength and determination to complete this journey and endure the long distance.

To Shahar, my love.

Table of Content

Introduction	1
I. The neural substrate of reading and its distinction from spoken language	9
Materials and Methods	12
Results	20
Discussion	27
II. Invariant neural responses to written and spoken language	30
Materials and Methods	31
Results	33
Discussion	36
III. Top-down attention modulation of neural response to written and spoken language	38
Materials and Methods	40
Results	52
Discussion	61
IV. Top-down attention modulation of inter-regional neural coupling of written and spoken language	64
Materials and Methods	67
Results	70
Discussion	75
General Summary	77
General Discussion	82
Table 1	91
Table 2	91
Table 3	92
References	94

Introduction

Until approximately five thousand 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 medium for communication that bridged geographical distance and time, therefore dramatically extending the information available to humans and their memory. This key pillar of modern civilization became possible by making the face-to-face spoken word interaction unnecessary for sharing knowledge.

Although writing systems allow a purely visual form of communication, this communication is still heavily rooted in spoken language. By design, in alphabets and syllabaries, individual written characters represent speech sounds. This connection is not just historical, but is also manifested within individual development, where reading typically follows the acquisition of spoken language. Reading proficiency is tightly linked to auditory ability to distinct phonemes, and can be improved with auditory practice (Wagner and Torgesen, 1987; National Reading Panel 2000). Moreover, speech in its natural form is highly intertwined with visual information from the speaker's face, like lip movement (McGurk and MacDonald, 1976; Campanella and Belin, 2007), which can establish neural scaffolding for cross-modal interactions in language processing. Thus, reading has long been thought to rely on the same neural processes that support spoken language (van Atteveldt et al., 2004).

However, it seems like that relationship between the two cognitive processes is more complicated: while some neural functions are common to both written and spoken language, others drastically diverge across the two mediums. When reading, the brain must first transform the sight of edges and lines into linguistic form (processes specific to the understanding of

written, not spoken, words; Dehaena et al., 2002) and ultimately into complex and abstract meaning—which is shared to some degree with the spoken word. Furthermore, a direct semantic understanding of written words is also possible without the ability to process sounds, as observed in deaf individuals (Gibson et al., 1970; Mayberry et al., 2010). Unlike the naturally learned comprehension of speech, literacy acquisition requires extensive training and is more susceptible to cognitive deficits, such as in ADHD and reading disorder (RD; American Psychiatric Association, 1994). This evidence suggests a more complex relationship between spoken and written language, with some neural components that are possibly unique to literacy.

How exactly does the human brain accommodate our newly developed skill of literacy with the evolutionarily older ability to comprehend speech? This dissertation is set to explore which neural processes are common or diverge across the auditory and visual language modalities.

Correspondence and distinction in comprehension of written and spoken language

Prior neuroimaging work has mapped the regions responsive to linguistic stimuli presented visually (writing) and auditorily (speech). Modality selective language processing was observed in the sensory cortices: the early auditory cortex responds to spoken (but not written) stimuli, whereas the visual cortex responds to written (but not spoken) stimuli. Because written words access the language system via the visual system, their comprehension requires sensory processing that is different from that required for understanding spoken words.

Within the visual system, extensive regions of the left ventral occipitotemporal cortex (also known as Visual Word Form Area) are specifically involved in reading (Cohen et al., 2000; Dehaene et al., 2001). Along this cortical region, more posterior areas are thought to be involved in visual feature extraction, and more anterior areas are involved in lexico-semantic processing of the whole word (Vinckier et al., 2007; Brem et al., 2010; Nosarti et al., 2010; Woollams et al.,

2010; Seghier and Price, 2011). While these areas are commonly agreed to be less involved in speech (Cohen et al., 2002; Booth et al., 2002a; 2002b), it is still a matter of debate whether their response properties are different for other non-linguistic visual stimuli (e.g. Vandenberghe et al., 1996; Moore and Price, 1999; Chee et al., 2000; Price and Devlin, 2003; Joseph et al., 2006).

Unlike its parallel in the visual system, no region in the auditory system was found to be exclusively activated by language. Despite the early identification of the left posterior superior temporal cortex (Démonet et al., 1992; Zatorre et al., 1992) and the anterior superior temporal gyrus (Binder et al., 2004; Obleser et al., 2006; Leaver and Rauschecker, 2010) as processing sites for the unique acoustic patterns of human speech, later studies controlling for auditory complexity and familiarity showed these regions to also process other non-speech auditory inputs (Wise et al., 2001; Narain et al., 2003; Specht and Reul, 2003; Thierry et al., 2003; Benson et al., 2006).

Further outside of the sensory cortex, text comprehension involves areas that are also active in speech comprehension and that therefore could be considered as invariant to the language modality. Modality-invariant activations were observed in widespread regions of the language system (Raij et al., 2000; Shaywitz et al., 2001; Booth et al., 2002b; Marinković, 2004; van Atteveldt et al., 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012). Specifically, a co-activation is common to the posterior superior temporal cortex, involved in grapheme to phoneme conversion and syntactic processing (Mesulam, 1998; Booth et al., 2002b; Friederici et al., 2009; Richardson et al., 2010); the inferior parietal cortex, involved in cross-modal integration of semantic features (Homae et al., 2002; Noppeney and Price, 2002; Schmithorst et al., 2006); and a subset of the inferior frontal cortex, involved in selecting and retrieving task-relevant semantics (Kang et al., 1999; Friederici et al.,

2000; Moro et al., 2001). Thus, prior studies have suggested a hierarchical model in which early sensory regions are modality-selective, but the written and spoken systems converge so that modality invariance increases toward higher-order semantic language regions (Chee et al., 1999).

The modality-invariant cortical areas could serve as multiple points of convergence of the linguistic modalities, as a result of the coextending routes of written word processing. Previous studies have revealed (at least) two parallel pathways leaving the occipital cortex for further high-level processing of text, each of them converging into general amodal steps in language comprehension (Jobard and Crivello, 2003). The phonological decoding route, which converts the sublexical graphemes into phonemes, links the superior temporal and ventral inferior parietal cortices to the dorsal precentral gyrus (Paulesu et al., 2000; Jobard and Crivello, 2003; McDermott et al., 2003; Price and Mechelli, 2005; Booth et al., 2006). The lexico-semantic reading route, which retrieves phonology directly from the whole written word or from its semantic properties, integrates the left ventral occipitotemporal cortex with the left ventral inferior frontal gyrus (Herbster et al., 1997; Jobard and Crivello, 2003; Mechelli et al., 2005; Graves et al., 2010). While unskilled readers are likely to initially rely more on the phonological route to decode unfamiliar words, the extensive experience of proficient readers will increase the use of the more direct lexico-semantic route in addition to the phonological one (Bolger et al., 2008a; 2008b; Levy et al., 2009).

Correspondence and distinction in top-down modulation of written and spoken language

Once reading proficiency is achieved, both phonological representation and the advanced semantic meaning of written words can be effortlessly processed, similar to the readily acquired speech. In fact, behavioral experiments have famously showed the semantic content of written words to have been automatically extracted, even when the text was not relevant to the

participants' current goals (Stroop, 1935; Catena et al., 2002). Such an automatic high-level processing effect is akin to the semantic processing of unattended speech, which has been studied extensively (Moray, 1959; Treisman, 1960; Lewis, 1970; Treisman et al., 1974; Elliott et al., 1998). The extent to which unattended language (and other types of stimuli) is processed is still debatable, to a large extent because of the variety of experimental designs used to manipulate top-down attention away from language and the disagreement in defining and assessing levels of attention and consciousness (Lachter et al., 2004; Kouider and Dehaene, 2007). Nevertheless, it is clear that the semantic content of irrelevant linguistic information that cannot be explicitly reported by participants is still implicitly processed and influences their behavior.

In addition to behavior, brain activity was also used as an indirect measurement for assessing automatic high-level processing of irrelevant linguistic input, adding more contradictory evidence to the debate. Some (but not all) studies, using both electrophysiology and imaging methods, showed processing of irrelevant speech in the modality-specific auditory cortex and modality-invariant linguistic areas (e.g. Jancke et al., 2001; Pulvermüller and Shtyrov, 2006; Sabri et al., 2008; Kouider et al., 2010; Ding and Simon, 2012; Rama et al., 2012; Golumbic et al., 2013; Horton et al., 2013). Automatic neural processing of written words was observed in some (but not all) studies in modality-specific visual cortex (including VWFA) and invariant linguistic areas (e.g. Fuentes et al., 1994; Price et al., 1996b; Rees et al., 1999; Dehaene et al., 2001; Ruz et al., 2005; Kouider and Dehaene, 2007).

Overall, the results described above have suggested a set of common neural substrates and similar top-down modulation effects for both written and spoken language modalities. However, the neuroimaging evidence that attempted to map these hierarchical influences and the

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 and ongoing attention in complex environments (Lerner et al., 2011; Ben-Yakov et al., 2012). Thus, they may underestimate the responses in high-order regions to spoken and written language, or to attended and unattended language, as well as their overlap.

Second, the demonstration of spatially overlapping neural responses is not strong evidence for modality or attention-level invariance. An aggregate activation may be observed in both modalities or both attended and unattended states 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 or attention-level invariance, a stronger form of invariance is indicated when a region responds with the same temporal response profile to attended and unattended spoken and written forms of the linguistic input.

This dissertation explores commonalities in neural processing mechanisms across written and spoken language by comparing temporally extended responses to real-life linguistic stimuli. Participants have been presented with auditory and written versions of engaging stories, which allows the capturing of neural activity underlying spoken and written language across the entire processing hierarchy in its full real-world complexity. Real-life linguistic stimuli have been shown to evoke highly reliable and selective responses, even in regions that show little response modulation to isolated letters, words, or sentences (Lerner et al., 2011; Silbert et al., 2014). Inter-subject correlation analysis (ISC; Hasson et al., 2009) provides an ideal tool for measuring the

similarity of the response time course to complex spoken and written stimuli.

In Chapter 1, the regions involved in text processing are mapped, and those regions unique to reading and not to speaking have been identified. In order to map sensory, linguistic, and extra-linguistic areas involved in text comprehension, participants were presented a real-life temporally extended narrative word by word in the center of a screen inside a functional magnetic resonance imaging (fMRI) scanner. This allowed for an assessment of the reliability of the response to the text across the readers. Previous studies (Lerner et al., 2011; Honey et al., 2012) of real-world speech comprehension have shown that a large portion of the cortex displayed reliable and selective responses to spoken stories, and that these responses were shared across subjects. By comparing levels of response reliability across the modalities, we can identify modality-specific areas.

In Chapter 2, the regions that show functional similarity between reading and listening to real-world language have been mapped by measuring the responses to the same written or spoken story. Calculating response similarities between the reading system and listening system from Chapter 1 allowed for identifying regions that had similar dynamic neural responses to language across modalities and were modality-invariant.

After identifying modality-specific and invariant cortical regions, Chapter 3 explores the extent to which top-down attention modulates responses to text and speech along their processing hierarchies. Written and spoken distinct stories were presented simultaneously while participants were asked to attend one story while ignoring another. By comparing a region's story-specific response (from a separate unimodal condition) to responses from the complex conditions of simultaneous presentation, we can identify areas that process the story when its content is attended or not.

In Chapter 4, the neural pathways through which a written and spoken attended and unattended content is distributed in the brain are identified. The neural communication between brain regions is examined using Inter-subject Functional Correlation (ISFC) analysis. In this novel methodological approach (Simony et al., 2016), network correlations are measured across brains in order to isolate stimulus-dependent processes. By isolating stimulus-induced correlation patterns, ISFC provides a window into large-scale brain dynamics supporting information processing. This innovative method allows determining, for the first time, the role of large-scale functional visual and auditory neural pathways in active language processing, and their interactions.

In summation, this dissertation sheds light on our unique ability to process language from multiple sensory sources. The findings attest to the remarkable plasticity of the human brain, which learns within a few years of training to extract linguistic information—auditory in origin—from purely visual signs, and to flexibly negotiate these different types of linguistic input. The new knowledge gained by this research work may allow us to better understand and treat reading deficits, as well as point us toward more effective pedagogical strategies to achieve literacy.

I) The neural substrate of reading and its distinction from spoken language

With the increasing importance of reading in the human culture, a great amount of research in the past few decades has been dedicated to mapping the brain regions involved in processing written language. While visual word recognition was originally localized in patients with brain damage in the left angular gyrus (Dejerine, 1891), the last 20 years have clearly established that written language understanding relies on activation in multiple brain areas. The locations of these regions have been identified with functional neuroimaging techniques such as positron emission topography (PET) and fMRI due to their superior spatial resolution. Because of variability in reading pace across participants and researchers' attempts to control for the sub-components of written language, text was most often studied by examining the averaged neural response to isolated graphemes, words, or sentences.

By and large, the circuit found to be involved in the comprehension of visually presented language was shown to be comprised of initial perceptual processing in the occipital lobe (Cohen et al., 2000; Dehaene et al., 2001) and widespread linguistic activation in temporal, frontal, and parietal areas (Crozier et al., 1999; Keller et al., 2001). By its nature, the visual cortex is involved in sensory processing of written inputs, although a region in the left temporo-ventral lobe (also known as the VWFA) was claimed to have a unique role in extracting the semantic content of single words. This early semantic processing is thought to be unique to the visual language modality, demonstrating higher average BOLD activation in response to written than spoken words (Cohen et al., 2002; Booth et al., 2002a; 2002b). This is in contrast to other stages of semantic processing in the temporal and frontal regions, as well as lexical processing, which

are thought to be equally active in reading as in speech comprehension (Raij et al., 2000; Shaywitz et al., 2001; Booth et al., 2002b; Marinković, 2004; van Atteveldt et al., 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012).

Unlike the visual system, the auditory cortex does not seem to have a parallel semantic region for speech; it seems to only be involved in the sensory processing of auditory language (Wise et al., 2001; Narain et al., 2003; Specht and Reul, 2003; Thierry et al., 2003; Benson et al., 2006).

Hence, prior studies have suggested a hierarchical model in which sensory regions are modality-selective, but the written and spoken systems quickly converge for further amodal high-order semantic processing in associative cortices.

Nevertheless, traditional neuroimaging paradigms do not have the ability to map the full set of regions involved in literacy. Recent studies have shown that aggregated activations to constrained linguistic stimuli could underestimate neural responses in high-order regions to written (and spoken) language (Lerner et al., 2011; Ben-Yakov et al., 2012). For example, Ben-Yakov et al. (2012) found that some regions produce a response with a temporal profile specific to each sentence and the context in which it is presented. Event-related averaging was found to often obscure this information, which makes up a substantial proportion of the responses to real-life stimuli. In addition, abstract linguistic content such as semantic meaning often manifests itself in complex stimuli spanning long periods of time.

Assessing response in neurons that are continually processing the unfolding linguistic information calls for inspecting the entire temporal profile of their response, not just its overall magnitude. Our approach was therefore based on measuring the reliability of response profiles over time. Response reliability offers a measurement of how effective a stimulus is in driving activity, and is complementary to the more common measure of response amplitude. Previous

studies using this method showed many brain regions to be strongly correlated in their activity within and across individuals when listening to a spoken narrative or watching a movie (e.g. Hasson et al., 2004; Golland et al., 2006; Wilson et al., 2007; Lerner et al., 2011).

Written and spoken language can differ not just in their sensory qualities, but also in their inherent cognitive demands. The behavior and practice involved in comprehension of text might require not just extraction of different sensory information, but also a different set of cognitive control processes (e.g., working and long term memory, attention, and even motion). For example, spatially directing attention is more important in reading than in listening and therefore might always evoke response in related posterior parietal areas (Brannan and Williams, 1987; Brunn and Farah, 1991; Hari and Renvall, 2001). The sensory and cognitive components unique to reading or listening are tightly entangled in each of these linguistic tasks, and are mostly inseparable. As a result, in this experiment the term “modality” refers to both “sensory modality” (auditory vs. visual) and “task modality” (listening vs. reading).

In the current experiment, 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. Each subject in this study either heard or viewed a spoken or a written version of the same continuous narrative while undergoing an fMRI. A rapid serial visual presentation (RSVP) of the written story, corresponding to the rhythm of the spoken content, enabled a reliable delivery across participants and correlation of their response profile. Subjects were instructed to attend to the details of the narrative, and a post-scan questionnaire assessed their comprehension and engagement. This design allowed for the mapping of written and spoken responses and the identification of modality-selective responses in areas that responded more reliably across subjects to the written (or spoken) narrative.

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 20-channel head coil. For functional scans, images were acquired using a T2*-weighted echo planar 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 $3 \times 3 \text{ mm}^2$ [field of view (FOV), $192 \times 192 \text{ mm}^2$].

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 mm³ resolution; FOV, 256 mm²). 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 (Figure 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.

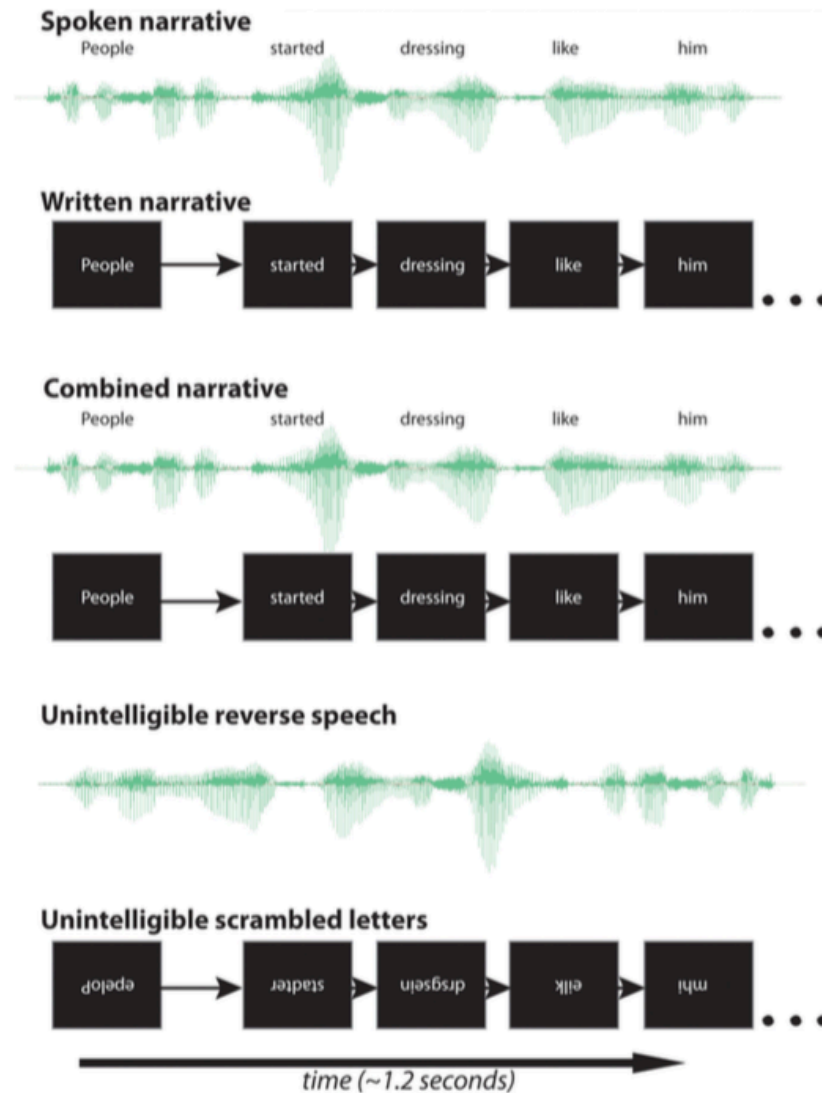


Figure 1. A 1.2 s segment of the 7 min narrative stimulus. While undergoing fMRI, subjects were exposed to one of two narrative presentation modes: a spoken version or a written version. Subjects were also exposed to a spoken and written control conditions. In the spoken control, the narrative was played 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.

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 (i.e., spoken narrative, written narrative, combined 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; 2009; 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. In a standard GLM analysis, experimenters usually assume prototypical response profiles 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 within each condition (i.e., within the reading group and within the listening group).

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). Family wise 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, 2001). 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 $R^* 0.16$.

To identify areas that show increase in response reliability for one condition over the other, a t test (0.05) was performed within each voxel that exceeded the threshold in at least one of the inspected conditions (see Fig. 4A). Thus, the t test was performed by comparing the correlation values of subjects from the spoken condition $\{r_j, r_{j1} \dots r_n\}$ to the correlation values of subjects from the written condition $\{r_j, r_{j1} \dots r_n\}$, 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 $10 \times 10 \times 10 \text{ mm}^3$ cubes, located on the peaks of the audio regression in each hemisphere. (2) A set of independent ROIs (see Figs. 4, 5) was defined based on an inter-subject 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 $\{\bar{r}_j, \bar{r}_{j+1} \dots \bar{r}_n\}$ to the mean correlations value of all voxels within the ROI from all subjects at the written condition $\{\bar{r}_j, \bar{r}_{j+1} \dots \bar{r}_n\}$.

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

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. 2A). The subjective level of difficulty recalling the narrative was low and equal across the two conditions ($t_{(13.99)} = -1$, $p = 0.33$; Fig. 2B). The subjective level of (13.99) engagement was high and also equal across conditions ($t_{(8.57)} = -1.41$, $p = 0.19$; Fig. 2C). 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.

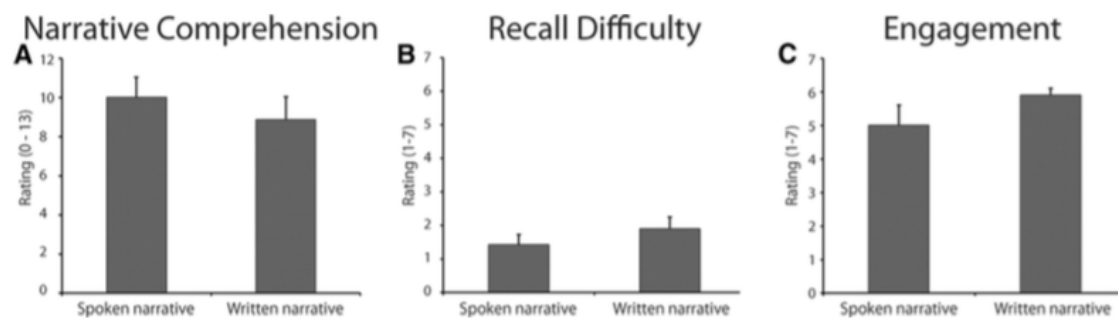


Figure 2. Behavioral measures of comprehension and engagement did not differ between the spoken and written conditions. **A**, Comparable levels of narrative comprehension were measured across the two modes of narrative presentation. **B**, Difficulty in recalling the narrative was equal across all conditions.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

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.

Inter-subject 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 extra-linguistic areas (Fig. 3A). 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, extra-linguistic regions, which seem to be involved in processing the narrative and the social content of a 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).

Inter-subject correlation in the written condition

Computing the inter-SC within the written condition (Fig. 3B) revealed reliable responses across subjects in the occipital visual cortex, as well as in many of the linguistic and extra-linguistic

areas observed in the spoken condition (Fig. 3C). Visual areas included early medial sensory cortices as well as higher visual areas in the ventral occipitotemporal cortex bilaterally. In addition, reliable response was also exhibited in other cortical areas, including the pSTG and pSTS, anterior superior temporal gyrus (aSTG), angular gyrus, and IFG, all in the language network, and extra-linguistic areas including the precuneus, the aIPL, the PCC, the dmPFC, DLPFC, and the orbitofrontal cortex (Table 2).

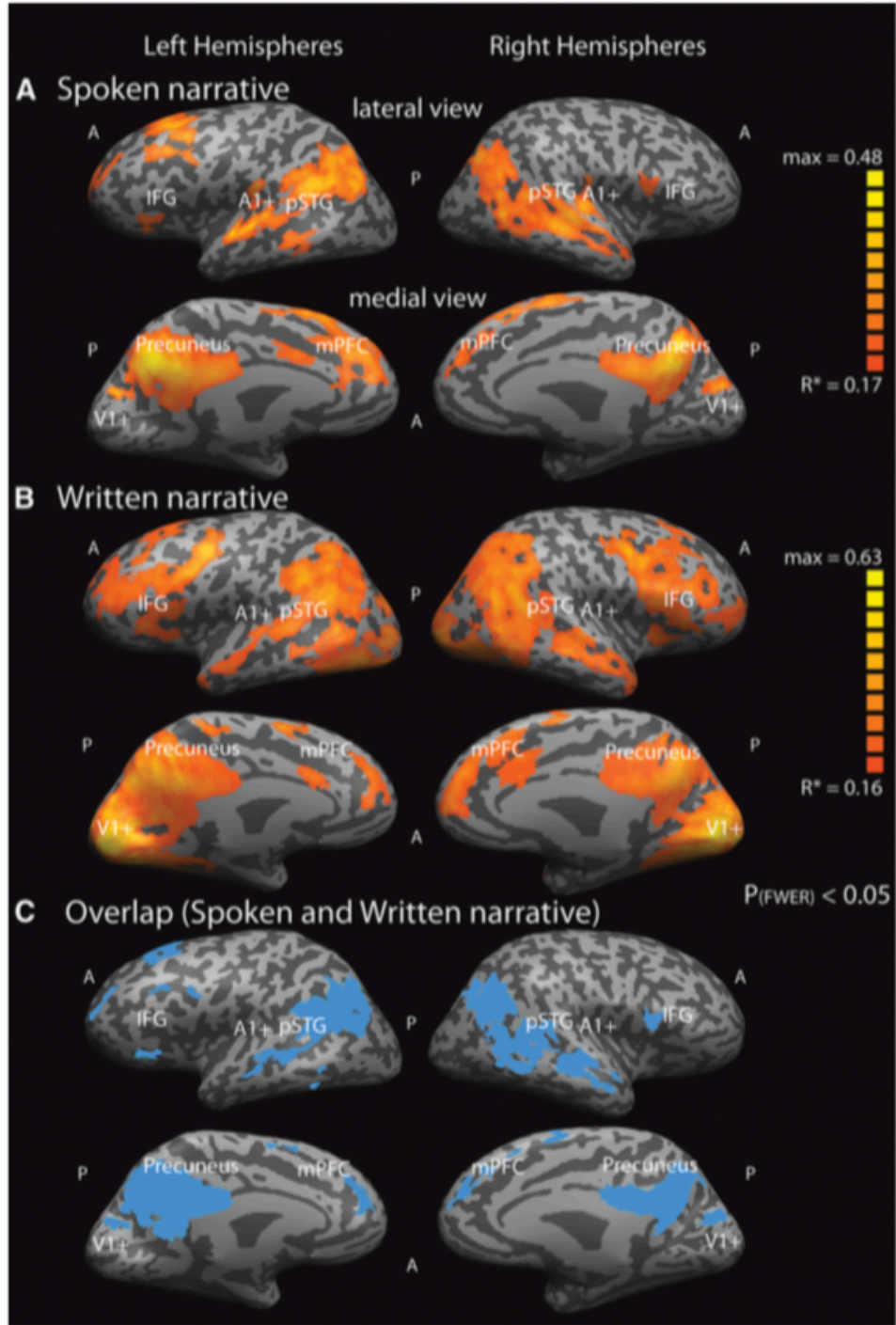


Figure 3. Reliability of brain response within the spoken and the written conditions. The fMRI BOLD time course in each voxel was correlated across subjects to produce a map of inter-SC within each presentation mode. **A**, **B**, The surface maps show the areas exhibiting reliable responses for (**A**)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

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 triangularis, the right orbitofrontal cortex, and posterior region within the right dmPFC (Fig. 4A, 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. 4A, 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. 4B). 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 ($t_{(16)} = 8.08$, $p < 0.0001$). Moreover, responses in A1+ were reliable even when the speech was unintelligible (played backward), 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.

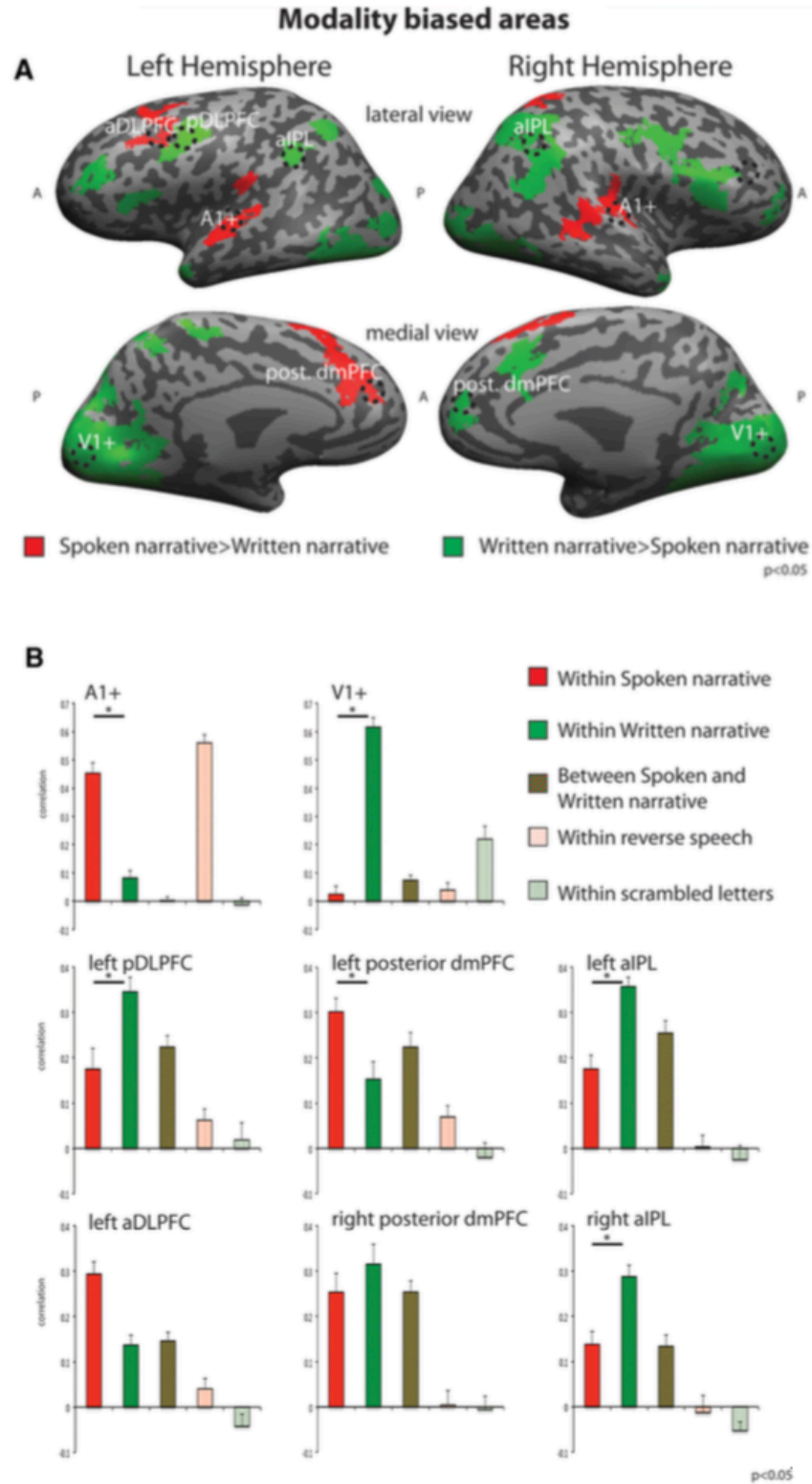


Figure 4. Brain regions exhibiting modality-selective bias to the spoken or written narratives. *A*, *t* tests between the spoken and written conditions within each voxel revealed areas that exhibit more reliable responses in the spoken condition (red) and other areas that exhibit more reliable responses 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.

Discussion

The analysis of temporally extended neural responses to complex written content revealed reliable responses across subjects in the visual cortex, as well as in many linguistic and extra-linguistic areas. The involvement of linguistic regions such as the pSTG, posterior middle temporal gyrus (pMTG), and the IFG is consistent with previous imaging studies that have suggested their role in semantic and lexical processing of text (see Price, 2012). The reliable response also included frontal and parietal extra-linguistic regions, known for their involvement in a variety of high-order cognitive processes and processing of the semantic and social aspects of a narrative (Fletcher et al., 1995; Xu et al., 2005; Ferstl et al., 2008; Lerner et al., 2011).

Given the uncommon reading task, where words appear in the middle of the screen at an externally controlled pace, our design is not suitable for examining reading processes that involve active eye movements. For instance, when reading from a page, the eye must execute a series of accurate saccadic movements along the text (Rayner, 1978), which are manifested in neural activation in the frontal eye field (FEF; Michael et al., 2001; Choi et al., 2014). While reading with the RSVP technique has been previously shown to evoke a level of difficulty in language comprehension equal to that of a page format (Mills and Weldon, 1987), the lack of eye movements in this design is likely to prevent the mapping of motor activation in the FEF, typical to most readers (Choi and Henderson, 2015).

By contrasting responses to written versus spoken content, we were able to reveal modality-

selective responses, which were not restricted to early visual and auditory cortices, but were also observed in parietal and frontal cortices (Fig. 4). Although the sensory regions are expected to exhibit modality-selectivity, the observation of selective responses in parietal and frontal areas was surprising in light of their suggested amodal control functions (Mesulam, 1998; Chee et al., 1999).

In particular, a greater reliability was observed 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 the 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 the anterior ventrolateral prefrontal cortex (VLPFC) and for vocalizations in the 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 the 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., auditory 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 responded reliably to the meaningful linguistic stimuli and not to unintelligible scrambled letters or sounds (Fig. 4).

II) Invariant neural responses to written and spoken language

Although writing systems allow a purely visual form of communication, they are still heavily rooted in spoken language. Reading proficiency is tightly linked to auditory skills, such as phonological awareness, and could be improved with further auditory practice (Wagner and Torgesen, 1987; Lovett et al., 1994; National Reading Panel, 2000; Ehri et al., 2001). Moreover, speech in its natural form is highly intertwined with visual information from speaker's face, like lip movement (McGurk and MacDonald, 1976; Campanella and Belin, 2007), which can establish neural scaffolding for cross-modal interactions in language processing (Dehaene, 2005; Dehaene and Cohen, 2007). Thus, reading has long been thought to rely on the same neural processes that support spoken language (van Atteveldt et al., 2004). We next turn to explore regions that show modality-invariant responses when processing complex language.

Using fMRI, modality-invariant activations were observed in widespread language systems, including the posterior superior temporal cortex, inferior parietal cortex, and subsets of the inferior frontal cortex (Shaywitz et al., 2001; Booth et al., 2002b; Marinković, 2004; van Atteveldt et al., 2004; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012). These regions are known for their relatively high-level role that follows the initial perceptual processing, which includes extracting semantic and lexical content from spoken and written language, and grapheme to phoneme conversion.

Regions that show “modality-invariant” responses were defined by previous neuroimaging studies as those that responded to both written and spoken stimuli, usually in isolated linguistic format (e.g., single words and sentences). For example, Vagharchakian et al. (2012) identified some linguistic regions in the left hemisphere as “modality-independent” based on overlapping

BOLD activation in response to isolated written and spoken sentences. However, spatially overlapping neural responses are 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, similar temporal response profiles to spoken and written forms of the same linguistic input would be stronger evidence of a region's modality-invariance.

Using the same data set from Chapter 1, temporal profiles of neural responses were compared across groups who read and listened to the same 7 min long story. ISC analysis (Hasson et al., 2010) is an ideal tool for measuring the similarity of the response time courses to these natural stimuli. This design allowed for the identification of modality-invariant responses in areas that demonstrated a reliable response across readers as well as across listeners in Chapter 1.

Materials and Methods

Subject

The subject population was the same as in Chapter 1.

MRI acquisition

Subjects were scanned in a 3T full-body scanner (Skyra, Siemens), using the same setup as in Chapter 1.

Stimuli and experimental design

To measure similarity in neural activity across the language modalities, we used the measured brain activity of listeners and readers of a real-life story, as in Chapter 1.

Data analysis

fMRI data were preprocessed and the mean of the white matter was projected out from the BOLD signal in the same manner as in Chapter 1.

Inter-SC maps were produced across the spoken and written conditions. The inter-SC maps provide a measure of the similarity of brain responses across the two conditions by quantifying the correlation of the time course of BOLD activity across subjects listening to the spoken narrative and reading the same written narrative.

For each voxel, 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. 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.

Bootstrapping by phase-randomization: The statistical likelihood of each observed correlation was assessed using a bootstrapping procedure based on phase randomization, as in Chapter 1. For each randomly phase-scrambled dataset, we computed inter-SC (R) for all voxels in the exact same manner as the empirical correlation map described above, i.e., by calculating the Pearson correlation between that voxel's BOLD time course in one individual from one group

and the average of that voxel's BOLD time courses in the other group of subjects. To correct for multiple comparisons, we selected the highest inter-SC value from the null distribution of all voxels in a given iteration, as in Chapter 1, producing a threshold of $\tilde{R}^* = 0.17$.

ROI analysis: We used the same ROIs as in Chapter 1.

Results

Modality-invariant responses

Spatial overlap of regions responsive to spoken and written narratives

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. 3C). 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. 5A), followed by whole brain analysis (Fig. 5B), and ROI analysis (Fig. 5C). Written and spoken narratives clearly evoked similar mean response time courses in the pSTG and precuneus (Fig. 5A). Equally modality-invariant responses were observed in the angular gyrus, IFG, anterior dmPFC, and left DLPFC (Fig. 5B; Table 2). The results of the voxelwise analysis were reproduced in a group of independent linguistic ROIs, such as the left IFG, the left pSTG, and left and right angular gyrus, and extra-linguistic ROIs such as the precuneus (Fig. 5C). 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 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 ($p_{(FWER)} < 0.05$). 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. 4B, 5C). These data rule out the possibility that regular stimulus onsets could have elicited the cross-modally reliable responses.

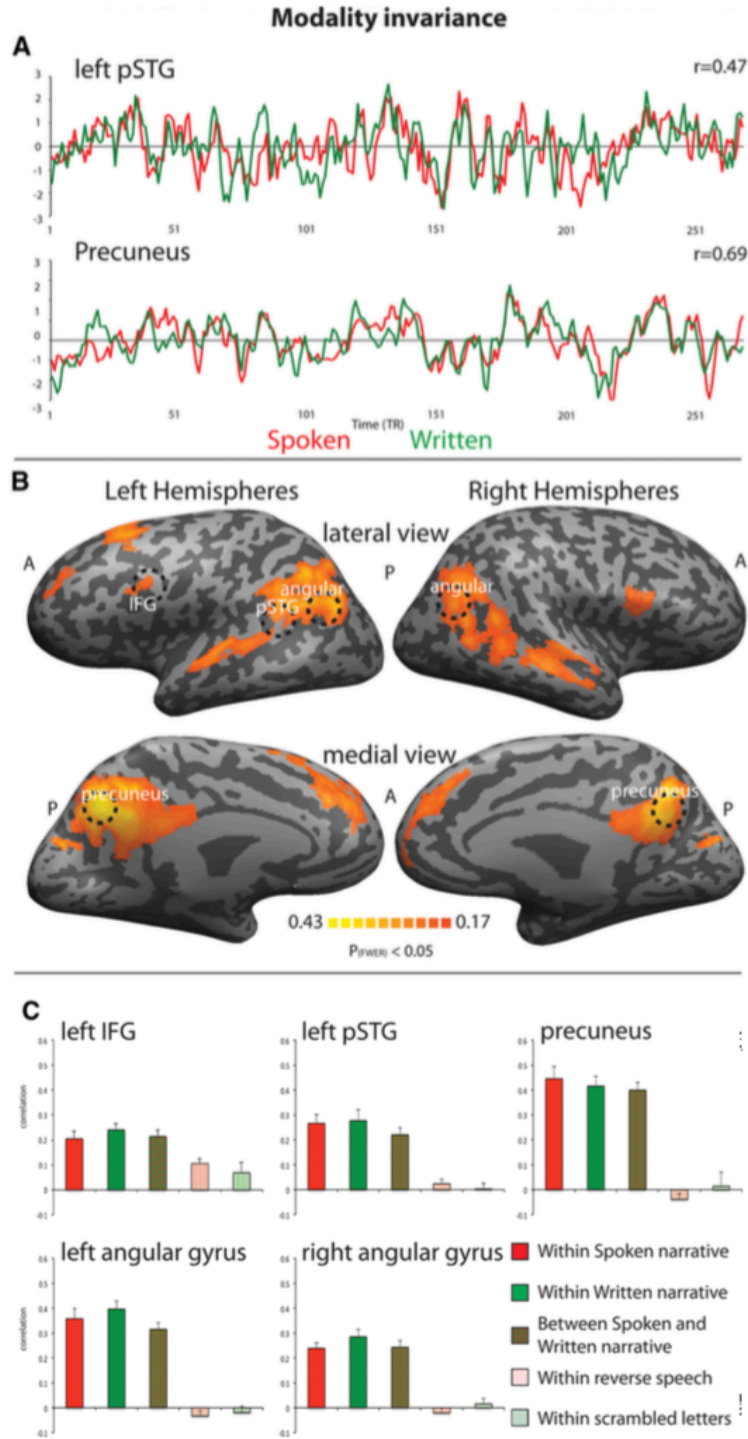


Figure 5. Brain regions exhibiting modality-invariant responses to the spoken and the written narratives. **A**, The average time courses of the responses in the left pSTG and the precuneus evoked by the written (green) and spoken (red) narrative. **B**, The fMRI BOLD time course in each area was correlated 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.

Discussion

Analysis of the temporally extended neural responses revealed robust modality-invariant response time 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 also process that information in a similar fashion across modalities.

The spoken and written narratives we presented have little in common in terms of their low-level sensory properties. This fit well with the observation of modality-selective response patterns in early visual and auditory areas (Chapter 1, Fig. 4B). Nevertheless, both forms conveyed essentially the same meaning, as verified by our comprehension test (Chapter 1, 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., 2002b; 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 the 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 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.

III) Top-down attention modulation of neural response to written and spoken language

Once mastered, reading becomes second nature to most people. Nevertheless, in contrast to speech, achieving literacy requires an active effort and is possible only when some cognitive control processes (such as attention and working memory) are adequately developed in the learning student (Daneman and Blennerhassett, 1984). The extensive practice involved in acquiring the skill of reading decreases the cognitive load for comprehension of written content (Yeung et al., 1997), and so despite its initial developmental disadvantage, the task of reading becomes a strongly automatic process, similar to speech comprehension. To what extent are written and spoken content automatically processed in different areas of the brain?

During a given act of visual or auditory perception, attentional focus may act to exclude irrelevant items from the perceived environment while emphasizing relevant ones. Top-down focus of attention is thought to drive enhancement of target-related activation in brain areas encoding task-relevant information (Luck et al., 1997; Kastner et al., 1999), prioritizing it over many distractors competing for control of behavior (Corbetta et al., 1991; Chawla et al., 1999). While the idea of relative enhancement of relevant information has been widely accepted, the fate of the ignored information is under a considerable debate (Lavie and Tsal, 1994; Driver, 2001).

The question of whether ignored information might undergo high-level processing, at least to some degree, has been extensively investigated in written and spoken word processing. Although several behavioral studies have suggested that attention is needed for some form of word encoding (Fuentes et al., 1994; Besner and Stolz, 1999; Naccache et al., 2002; Lachter et al.,

2004), many others have indicated that unattended words access lexical and semantic analyses automatically (Stroop, 1935; Moray, 1959; Treisman, 1960; Lewis, 1970; Treisman et al., 1974; Fuentes and Ortells, 1993; Elliott et al., 1998; Catena et al., 2002).

Brain activity was also used as an indirect measure for assessing automatic high-level processing of irrelevant spoken and written language. Some neuroimaging studies have shown automatic linguistic processing of speech in the inferior frontal gyrus, planum temporale, and the superior temporal cortex (Jancke et al., 2001; Rama et al., 2012), and automatic linguistic processing of text in the left fusiform gyrus, the posterior temporal cortex, inferior parietal cortex, and frontal regions (Price et al., 1996a; Dehaene et al., 2001; Peng et al., 2004). Nevertheless, other studies have found no significant BOLD activity in high-order regions when irrelevant spoken and written language was unattended (Rees et al., 1999; Sabri et al., 2008; Kouider et al., 2010).

The neuroimaging evidence supporting automatic processing of irrelevant language at the same areas as fully attended language can be criticized in the same manner described in Chapter 2: First, spatial overlap between averaged responses to relevant and irrelevant stimuli does not guarantee identical input processing (Dinstein et al., 2007; Ben-Yakov et al., 2012; Honey et al., 2012). Second, isolated words and sentences cannot map the full set of regions engaged in automatic processing of real-life language (Lerner et al., 2011; Ben-Yakov et al., 2012). A more precise indication for automatic processing within a region is its similar temporal response profile to the same linguistic input across different states of attention.

Furthermore, in the majority of prior studies of attention to language (e.g. Price et al., 1996a; Banich et al., 2000; Pulvermüller and Shtyrov, 2006), task-relevant and task-irrelevant information were delivered within the same sensory channel, such that the irrelevant information constituted a feature within the attended stimulus. The use of an intra-modular design might not

show conclusively to what extent unattended linguistic information is processed. This study adopted a bimodal selective attention paradigm, in which relevant information is presented in one modality and irrelevant information in another, therefore offering the potential to achieve a better separation of attentional focus.

Using fMRI, we recorded neural responses from participants who were simultaneously presented with two unrelated stories, one spoken and one written (presented one word at a time). One group of participants was asked to attend to the written story while ignoring the spoken one, while the other group did the reverse. Their neural data were compared to two unimodal control groups, who were exposed to only one of the stories and hence shared with the simultaneous groups the experience related to that particular story. A comparison of the temporal response profile of the simultaneous groups with the unimodal groups allows the characterization of each local response as goal-relevant or irrelevant.

Materials and Methods

Subjects

Seventy-four subjects successfully participated in one of the two main experimental conditions (attention to text and attention to speech), or in one or two of the control conditions (unimodal text and speech). Eighteen subjects were discarded from the analysis: 6 subjects due to head motions >3mm, 5 for closing their eyes during the story, 1 due to corrupted anatomical signal, 1 due to anomalous anatomy, 1 due to difficulties in hearing the stimulus, 1 due to missing behavioral results, 1 due to previous familiarity with the story, and 2 due to failure in the memory tests (see behavioral assessment). Additional subjects were scanned until data from 18

subjects were collected for the attention to text (14 females; ages 19-32), attention to speech (14 females; ages 18-24), unimodal text (13 females; ages 18-29), and unimodal speech (13 females; ages 18-30) conditions. Sixteen of the subjects participated in both the unimodal text and speech conditions. Another 19 subjects were scanned in a rest condition (8 females, ages 18-31), and 13 subjects were scanned while presented with a different audiovisual story (7 females, ages 19-26).

Procedures were approved by the Princeton University Committee on Activities Involving Human Subjects. All subjects were right-handed native English speakers, reported normal hearing and normal or corrected-to-normal vision, normal reading skills, and had not heard the two stories prior to the experiment. All subjects provided written informed consent.

MRI acquisition

Subjects were scanned in a 3T full-body scanner (Skyra, Siemens), using the same setup as in Chapters 1 and 2.

The text was projected using an LCD projector onto a rear-projection screen located in the magnet bore, and was viewed with an angled mirror. Stimuli were presented and synchronized with MRI data acquisition onset using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) for MATLAB.

Stimuli

The spoken language stimulus was a 15:03 min real-life story (“Slumlord” told by Jack Hitt, recorded live at “The Moth” storytelling event, New York City). The written language stimulus was a 15:03 min transcript of a different real-life story (“The Overview Effect” told by Richard Garriot), recorded at the same live storytelling performance. The content of the two stories was not related.

In the written stimulus, the words were individually presented in a white font in the center of a black screen in a rapid serial visual presentation. Out of the 2381 words of the story, most of the words (2093) were presented for the duration of 310 ms each. 189 of the resulting words were accompanied with punctuation and thus presented for 650 ms, and the remaining 99 words appeared at sentences' ends and thus presented for 1200 ms. These varied durations aimed at providing an easy reading experience for subjects (Castelhano and Muter, 2001). The width of the words ranged from 4 to 160 pixels, the height ranged from 11 to 20 pixels.

In addition to the written and/or spoken stimulus, all conditions contained a red fixation point (radius of 5.6 pixels, 30% transparency) at the center of the screen, juxtaposed over the center of the words of the written stimulus, if presented. 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.

The spoken and written stories were combined to create a simultaneous auditory and visual presentation in the two main experimental conditions (Fig. 6). In the two control conditions, either the spoken or written stories were presented on their own, in a unimodal fashion

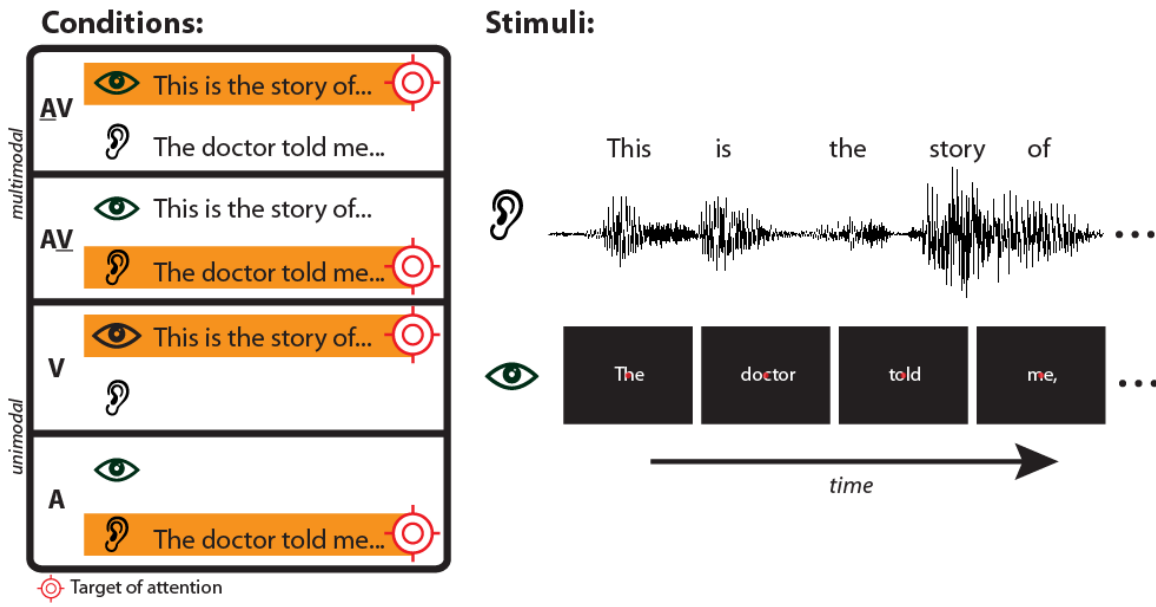


Figure 6. A segment from the 15 min spoken and written unrelated stories, presented as stimuli in four experimental conditions. While undergoing fMRI, two multimodal groups were exposed simultaneously to the visual (written) and the auditory (spoken) stories. One group (AV) attended the auditory story and the other (AV) attended the visual story. Two unimodal control groups were exposed to either the auditory story (A) or the visual story (V).

Experimental design

Participants were instructed to carefully attend and remember the details of the presented story and were informed they would receive a monetary bonus based on their performance in the subsequent memory test. In the two main conditions where the two stories were presented simultaneously, participants were instructed to ignore the distracting story: The attention to text group (AV) attended the written stimulus while ignoring the spoken one, while the attention to speech group (AV) did the opposite (Fig. 6). In the two additional control groups, participants were exposed to one of the stories alone: The unimodal text group (V) read the written story, and the unimodal speech group listened to the spoken story (A).

All participants who were to be exposed to written content in the experiment got to practice reading a sample of unrelated text presented in RSVP before the beginning of the trial. The volume of the auditory stimuli was adjusted individually for each subject to a comfortable and

clear level. When the spoken stimulus was attended in the attention to speech and unimodal speech groups, subjects were asked to fix their gaze toward the fixation point for the entire time. In the cases where subjects participated in both of the control conditions (89% of the subjects), the order of the two conditions was randomized.

Data analysis

Preprocessing: fMRI data was preprocessed in FSL (<http://fsl.fmrib.ox.ac.uk/fsl>), including slice time correction, motion correction, linear detrending, high-pass filtering (140 S cutoff), and coregistration and affine transformation of the functional volumes to a template brain (MNI). Functional images were resampled to 3 mm isotropic voxels for all analyses. All calculations were performed in volume space. Projections onto a cortical surface for visualization were performed, as a final step, with NeuroElf (<http://neuroelf.net>).

Inter-subject correlation analysis: ISC maps were produced across conditions (e.g., attention to speech group vs. unimodal speech control group). The ISC maps provide a measure of the similarity of brain responses between two different conditions by quantifying the correlation of the time course of BOLD activity between each subject from one group and the averaged activity of all subjects in the other group (Hasson et al., 2004; Honey et al., 2012).

For each voxel, ISC between two conditions is calculated as an average correlation:

$$\tilde{R} = \frac{1}{N} \sum_{j=1}^N \tilde{r}_j$$

where the individual $\tilde{r}_j$ are the Pearson correlations of that voxel's BOLD time courses of the j 'th individual from the first group and the average of that voxel's BOLD time course of all individuals in the other group.

The ISC was performed across the following conditions: attention to speech group vs. unimodal speech group; attention to text group vs. unimodal text group; attention to text group vs. unimodal speech group; attention to speech group vs. unimodal text group; unimodal text group vs. unimodal speech group.

ISC within a condition is calculated as an average correlation:

$$R = \frac{1}{N} \sum_{j=1}^N r_j$$

where the individual $\tilde{r}_j$ are the correlations between the BOLD time course in one individual and the average of the BOLD time courses in the remaining individuals. This within condition ISC analysis was performed within the group presented with a coordinated audiovisual story (see ROIs analysis).

In a standard GLM analysis, experimenters usually assume a prototypical response profile for each specific stimulus. The ISC 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, when performed across conditions, the ISC method uses the averaged brain response to the content in one condition within a given brain area as a model to predict brain responses to different content presented in the other condition. When performed within a condition, the ISC method uses the subject's brain responses as a model to predict brain responses to the same content.

Bootstrapping and phase-randomization: The statistical likelihood of each observed correlation was assessed using a bootstrapping procedure based on phase randomization, as in Chapters 1 and 2. The thresholds for each condition were as follows: attention to speech group vs. unimodal speech group $\tilde{R}^* 0.073$; attention to text group vs. unimodal text group $\tilde{R}^* 0.066$; attention to text

group vs. unimodal speech group $\tilde{R}^* 0.062$; attention to speech group vs. unimodal text group $\tilde{R}^* 0.063$; unimodal text group vs. unimodal speech group $R^* 0.066$. The same procedure was performed on the correlation values within the group of participants presented with a coordinated audiovisual story (see ROI analysis). Their correlation threshold was computed between each individual and the average of the remaining participants in the group, producing a threshold of $R^* 0.15$.

Functional Connectivity (FC) analysis: The FC correlation matrix was calculated between all voxels within the brains of subjects from the resting group. The neural signals X_i measured from subject i , $i = 1, \dots, k$, are in the form of a $p \times n$ matrix that contains signals from p neural sources over n time points. All time courses were z-scored within subjects to zero mean and unit variance. Thus, the subject-based FC was calculated by the Pearson correlation between different voxels within a single subject and then averaged across all group members.

Intrinsic connectivity networks (ICNs) and Regions of Interest (ROI) analysis

First, the entire cortex was filtered to include only voxels that have been demonstrated before as showing a reliable response to written and spoken content. For this purpose, voxelwise ISC analysis was performed over a separate group of subjects exposed to a coordinated audiovisual presentation of the same story (“Pie-man”, see Chapter 1 for more stimulus related details). In the process of defining ROIs and intrinsic connectivity networks (ICNs), only voxels that demonstrated a significantly reliable response to the coordinated audiovisual linguistic stimulus were included.

Next, we mapped ICNs by clustering patterns of functional connectivity that were calculated in a separate group of resting participants. FC analysis was performed between all filtered voxels

within each of the 19 subjects and then averaged across all group members. K-mean clustering was performed over the FC correlations using kmeans function in MATLAB to extract six ICN. This procedure partitioned the voxel-based FC correlation matrix into k mutually exclusive clusters. Each cluster was defined by a set of N member voxels (each with an associated correlation vector) and by the centroid of the correlation vectors in the cluster. The iterative algorithm minimizes the sum of distances from each voxel (vector) to its cluster centroid, over all clusters. Each cluster contained functionally connected voxels that were grouped as a network. The extracted networks were labeled based on their anatomical identification as: Auditory, Visual, DMN-I, DMN-II, Attention, and Executive (Fig. 7).

The six ICNs were then split into regions of interest (ROIs). We started by identifying the early auditory and sensory cortices within the auditory and visual networks. Heschl's gyri were defined within the auditory network based on the probabilistic Harvard-Oxford Cortical Structural Atlas (Desikan et al., 2006). V1, V2, V3, V3a and hV4 were identified within the visual network using a published retinotopic probabilistic atlas (Wang and He, 2014). In cases where a voxel was included in several regions simultaneously, it was assigned to the region with the highest probability to avoid overlapping ROIs. Four voxels from V1 that did not overlap with the visual network were discarded.

In further defining ROIs in the rest of the networks, voxels were clustered using a parcellation approach based on local connectivity patterns (Baldassano et al., 2015). This clustering method was performed within each network on the FC voxelwise correlations of the resting group, excluding the early sensory regions defined earlier. Each network was clustered into the minimal number of exclusive regions that would not extend 900 voxels each.

Finally, regions that were not spatially continuous (e.g., across hemispheres) were split. Along the process of defining ROIs, any region that included less than 10 voxels was dropped from further analysis, adding up to 271 dropped voxels in total.

Overall, this procedure yielded 61 ROIs: 15 in the visual network, 9 in the auditory network, 10 in the DMN-II, 6 in the DMN-I, 8 in the attention network, and 13 in the executive network (Table 3; Fig. 7).

Intrinsic connectivity networks and regions of interest

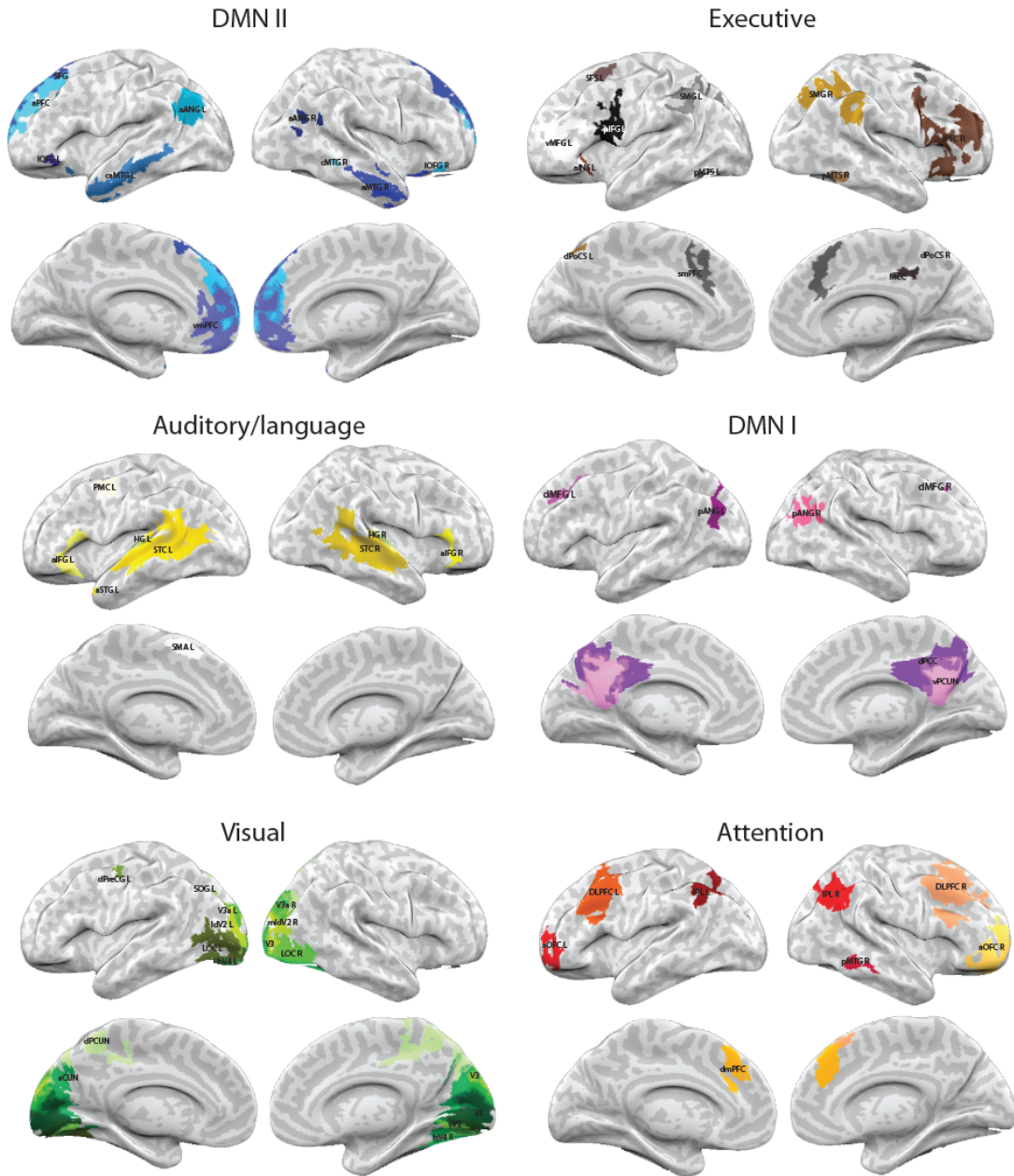


Figure 7. Cluster analysis of the FC covariance matrix during rest reveals networks organization structure, parceled into 61 regions of interest. K-means clustering of voxel-based FC covariance matrix during rest ($K=6$) reveals six intrinsic connectivity networks (ICNs). The cluster surrounding Heschl's gyri and the STC was labeled the auditory/language network (shades of yellow). The cluster surrounding the calcarine sulcus was labeled the visual network (shades of green). The DMN seems to split into two sub-networks labeled DMN-I (shades of pink) and DMN-II (shades of blue). Finally, the two dominantly frontoparietal networks were labeled as the attention (shades of red) and executive (shades of grey/brown) networks. Within each network, we identified early sensory regions of interest based on probabilistic maps. The rest of the voxels were clustered using a parcellation approach based on local connectivity patterns.

Behavioral assessment

Immediately following the scan, each subject's memory and engagement for the two stories was assessed using computerized questionnaires. Subjects in the simultaneous conditions were informed in advance about a memory test only for the story they were asked to attend, but were surprised after the scan with an additional memory test for the unattended story. Before answering the tests, it was explained to them that correct responses would win them a monetary bonus, regardless of the story. This procedure aimed at encouraging participants to answer the memory tests to the best of their abilities, regardless of whether they were originally instructed to attend them or not.

The order of the two story questionnaires was counterbalanced between subjects within each of the conditions. Subjects in the control conditions were also informed in advance of the memory tests for the one or two unimodal stories they attended (no unattended content presented). For the subjects who participated in both control conditions and therefore attended (separately) both stories, the order of the two questionnaires was pseudo-randomized within each of the conditions. The order of the questions within each story questionnaire was not randomized.

Attention: Subjects were asked to assess on a scale from 1 to 10 how much of the experiment time (0-100%) they spent attending the story.

Free recall: For the first memory test, subjects were asked to write down a summary of the narrative of the story, as detailed as possible. Three independent raters graded these written records (on a scale from 1 to 20), taking into account subjects' general story comprehension as well as memory for small details. Raters' grades were highly cohesive. The scores were z-scored within each rater, averaged across all three of them, and linearly transformed to positive values.

One subject's recall summary was not included in the analyses due to a technical problem in saving the data.

Multiple choice: In the second memory test, subjects were asked to answer forced-choice questions with four potential choices about the content of the story (24 questions for the written story and 25 for the spoken story). The percentage of correct answers was calculated for each subject.

Fill-in-the-blank: For the last memory test, subjects were given the story's transcript with missing words or phrases (84 in the written story and 77 in the spoken story) and were asked to fill in the missing text. Importantly, no information related to the missing words appeared in the previous multiple choice questions. Three independent raters graded each of the fill-ins (on a scale from 0 to 4), taking into account the subjects' general conceptual memory for the missing text as well as the accuracy of the wording. The percentage of the score received out of the total maximal score was calculated for each of the subjects and averaged across all three raters.

For each story questionnaire, one-way ANOVA tests ($\alpha = 0.05$) were conducted between groups to compare the effect of the task on (a) the quality of the rated free-recall test; (b) the success in the multiple choice test; and (c) the success in fill-in-the-blank memory tests. Post hoc comparisons using the Tukey-Kramer test assessed pairwise differences between all conditions. In addition, two-tailed t-tests ($\alpha = 0.05$) were conducted between the attending and not attending groups to compare the self-reported percentage of time spent attending to each of the stories.

An exclusion criterion was applied based on participants' multiple choice memory performances to screen for individuals who did not perform their tasks as expected. According to this criterion, any participants who deviated from the minimum requirement of 65% correct responses for the attended stimuli were excluded from both the behavioral and neural analysis. Two participants

from the group attending the spoken story were marked as outliers based on their low performance in the spoken story multiple choice test.

Eye tracking

Eye tracking was conducted using the iView X MRI-LR system (Sensomotoric Instruments [SMI]), sampling at 60Hz. The experimenter monitored participants' alertness and general direction of gaze via the eye tracking camera. Any participant who appeared not to be looking toward the monitor or who closed their eyes was excluded from all analyses. In addition, some participants were not included in the eye tracking analysis due to calibration problems or incomplete eye tracking data (i.e., missing more than 60% of the data), including 3 subjects from the attention to speech condition, 2 subjects from the attention to text condition, and 2 subjects from the unimodal control conditions.

Eye movements were measured to examine deviations in gaze from the written stimulus. For this purpose, the proportion of the gaze data that remained within the horizontal borders of the longest written word of the story (160 pixels long) was calculated for each subject. All participants remained within these borders for 95% of the recorded data samples, except for one participant in the attention to text condition (92% of data within the borders) and one participant in the unimodal speech control (91% of the data within the borders).

Results

Behavioral results

Participants in the simultaneous presentation groups (AV and AV) reported spending a bigger portion of the scan time attending the required story than they spent attending the story they were

asked to ignore. The subjective assessment of time spent attending was higher for the spoken story than for the written story when participants were asked to attend the speech ($t_{(34)} = 10.11$, $p < 0.0001$). The subjective assessment of time spent attending was higher for the written story than for the spoken story when participants were asked to attend the text ($t_{(34)} = 14.08$, $p < 0.0001$).

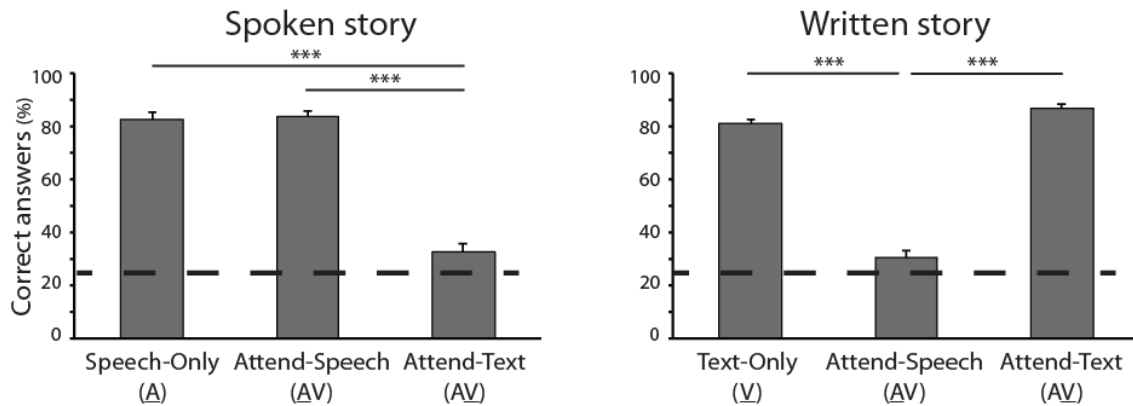
Participants remembered the stories they were instructed to attend better than the stories they were asked to ignore (Fig. 8). Participants who attended the spoken story performed better in all memory tests about the spoken content, compared to the group who attended the written content: The multiple-choice ($F_{(2,51)} = 161.8$, $p < 0.0001$), fill-in-the-blank ($F_{(2,51)} = 132.8$, $p < 0.0001$), and recall ($F_{(2,50)} = 58.9$, $p < 0.0001$) scores were higher in groups AV and A than in group AV. Participants who attended the written story performed better in all memory tests about the written content, compared to the group who attended the spoken one: The multiple-choice ($F_{(2,51)} = 288.7$, $p < 0.0001$), fill-in-the-blank ($F_{(2,51)} = 79.2$, $p < 0.0001$), and recall ($F_{(2,51)} = 107.8$, $p < 0.0001$) scores were higher in groups AV and V than in group AV.

Participants' memory for the attended stories was not influenced by the simultaneous presentation of another story. Memory scores were high and not significantly different across the group who attended that story while ignoring another and the control group that was exposed to that one story only. For the spoken story, memory performance in all tests was similar between group AV (Multi: $M = 83.6\%$ $SD = 5.8$; Fill-in: $M = 45.5\%$ $SD = 8.1$; Recall: $M = 3.39$ $SD = 0.8$) and the control group A (Multi: $M = 82.7$ $SD = 7.1$; Fill-in: $M = 44$ $SD = 12$; Recall: $M = 3.75$, $SD = 0.43$). For the written story, memory performance in all tests was similar between group AV (Multi: $M = 87\%$ $SD = 6.2$; Fill-in: $M = 47.2\%$ $SD = 11.7$; Recall: $M = 3.48$, $SD =$

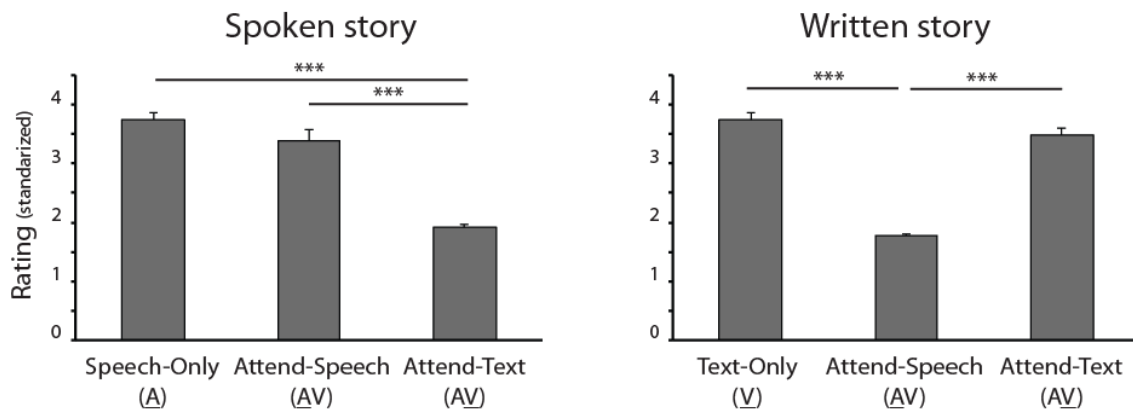
0.54) and the control group V (Multi: $M = 81$ $SD = 5.9$; Fill-in: $M = 48.5$ $SD = 10.7$; Recall: $M = 3.74$ $SD = 0.51$).

Overall, these results suggest that the research design successfully manipulated attention toward the desired story and away from the story that participants were instructed to ignore. Directing attention toward a story was associated with cognitive processing of its specific content, including increased narrative comprehension and long-term memory. In addition, the introduction of a potentially distracting story simultaneously did not seem to significantly influence participants' comprehension and memory of the attended story. These behavioral trends allowed us to continue and explore attentional modulation of neural responses.

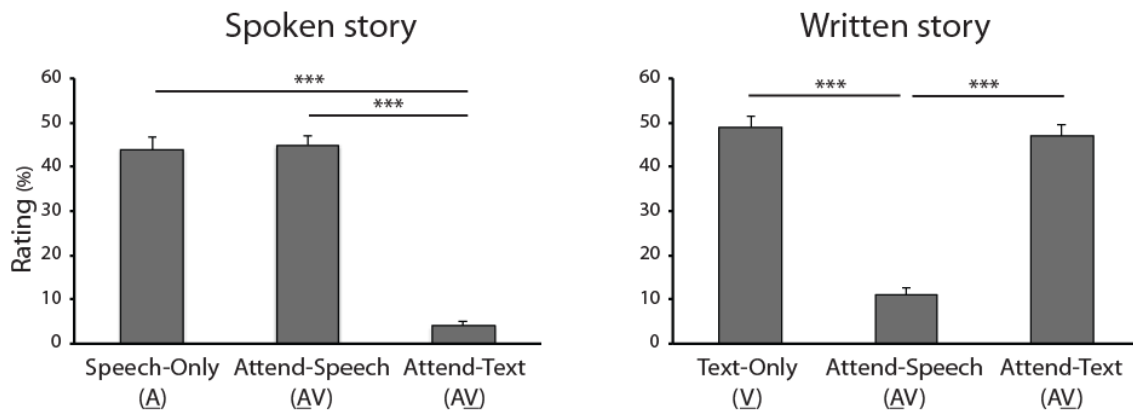
Multiple-choice



Free recall



Fill in the Blank



*** $p < 0.0001$

Figure 8. Behavioral measures of memory for spoken and written stories. Scores in the multiple choice, free recall, and fill-in-the-blank tests were higher in groups asked to attend the spoken/written story, compared to groups who ignored that content. Scores did not differ between groups who attended stories unimodally and groups who attended the same content while ignoring a distracting content (i.e., A vs. AV and V vs. AV). Values are means and error bars represent SEM across subjects.

Attentional modulation of goal-related responses within cortical regions

We identified brain areas that process the written story or the spoken story and how attention modulates activity in these areas. To do so, the response time course of the unimodal groups was correlated with the response time course of both attention groups (A to AV and AV; V to AV and AV) within every voxel. The statistical significance of this ISC analysis was assessed using a bootstrapping procedure based on phase randomization (see Materials and Methods).

Inter-subject correlations during spoken story

Neural responses during attention to the spoken story were highly similar to the response during the unimodal presentation of the spoken story. While attending the spoken story, significant ISC was measured in both early sensory cortices and a number of high-level regions that have been shown in Chapter 1 and previous work to process unimodal spoken stories (Lerner et al., 2011; Honey et al., 2012). Sensory areas included Heschl's gyri and higher auditory cortices in the STG (Romanski and Averbeck, 2009), but also, surprisingly, visual cortices, as early as V1 (independently defined by retinotopic probabilistic atlas; Wang and He, 2014). High-level regions include extended parts of the frontal, parietal, and temporal cortices, with strongest ISC values in the pSTG, angular gyrus, and pIFG, known for their linguistic processing (Hickok and Poeppel, 2007; Sahin et al., 2009; Price, 2010); the dPCC, mPFC, and IPL, recognized as parts of the default mode network and involved in processing the narrative and social content of the story (Xu et al., 2005; Ferstl et al., 2008; Lerner et al., 2011); and the ACC and DLPFC, recognized as part of attention and executive networks (Posner and Rothbart, 1998; Smith and Jonides, 1999; Duncan and Owen, 2000).

In contrast, when attention was directed toward the written story, only auditory and some linguistic areas showed similar activity to the unimodal presentation (Fig. 9B). These areas

included Heschl's gyri, the STC, aSTGL, IOFG, and the aIFG. High-level areas involved in the processing of temporally extended narratives did not show similar neural responses.

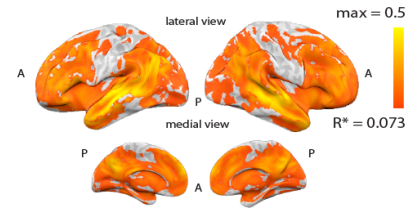
Inter-subject correlations during written story

While attending the written story, early visual cortices as well as higher-order cognitive regions showed similar responses to the processing of the written story alone (Fig. 9C). Sensory areas included the entire visual processing hierarchy, from V1 and up to lateral occipital cortices (Price and Devlin, 2011), while higher-order areas included temporal, parietal, and frontal regions, similar to those that responded to the attended spoken story, which are known to be involved in processing the linguistic and social content of stories. Most of these regions have been previously shown in Chapter 1 to have a reliable temporal response across subjects to a simple unimodal presentation of written stories.

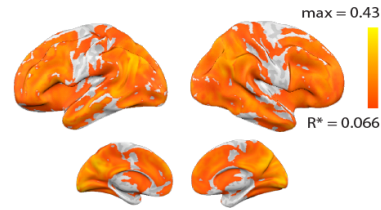
However, when attention was not directed toward the written story, only the visual areas and the posterior medial temporal gyrus (pMTG) continued to show responses characteristic to the processing of the text (Fig. 9D). The bilateral visual areas in the occipital cortex included V1, V2, V3, V3a, hV4, aCUN, and LOC. The ISC in a set of voxels in the pMTG was on average weaker than that in the visual cortex, but still significant, and more robust in the left hemisphere.

Processing with attention

A Speech processing (A vs. $\underline{AV}$)

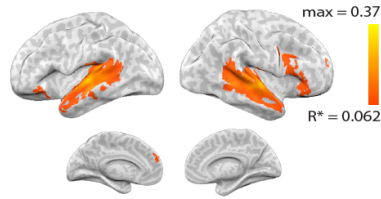


C Text processing (V vs. $\underline{AV}$)

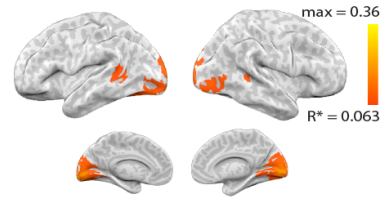


Processing without attention

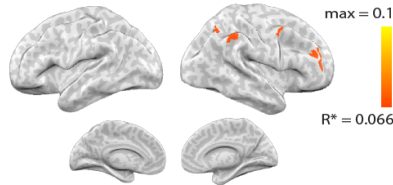
B Speech processing (A vs. $\underline{AV}$)



D Text processing (V vs. $\underline{AV}$)



E Baseline (V vs. A)



P(FWER) < 0.001

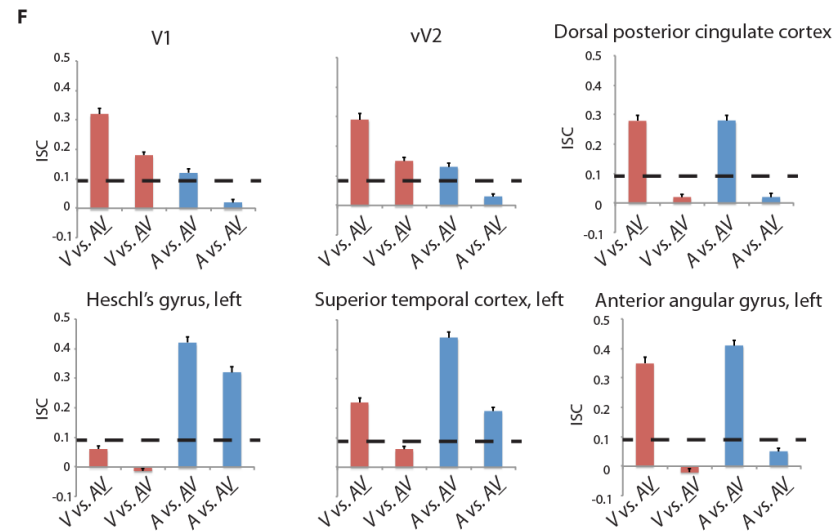


Figure 9. Brain regions exhibiting story-specific responses to spoken or written story in attended and unattended states. The fMRI BOLD time course in each voxel was correlated across the control group of unimodal listeners (A) and $\underline{A}$, the multisensory group who attended speech ($\underline{AV}$), or $\underline{B}$, the multisensory group who did not attend speech ($\underline{AV}$). The time courses were also correlated across the control group of unimodal readers (V) and $\underline{C}$, the multisensory group who attended text ($\underline{AV}$), or $\underline{D}$, the multisensory group who did not attend text ($\underline{AV}$). **E**, Regions exhibiting baseline similarity in responses between unimodal listeners (A) and readers (V). **F**, An example of independently defined regions of interest that

exhibited story-specific responses to spoken or written story in attended and unattended states. These regions included V1, ventral V2, dorsal posterior cingulate cortex, left Heschl's, left superior temporal cortex, and left anterior angular gyrus.

Quantifying modulation by attention in different cortical areas

Previous work has shown that attention enhances neural responses differently along the processing hierarchy (Jancke et al., 2001; Bluvas and Gentner, 2013; Golumbic et al., 2013; Wang and He, 2014). To measure such effects, we calculated the difference in inter-subject correlation levels across the attended and unattended states as a percentage of the attended state correlation in each voxel (i.e., $(corr(\underline{AV}, A) - corr(\underline{AV}, A))/corr(\underline{AV}, A)$ to measure effects of attention on processing speech and $(corr(\underline{AV}, V) - corr(\underline{AV}, V))/corr(\underline{AV}, V)$ to measure effects on processing text). Thus, the bigger the relative modulation in a voxel was, the more its involvement in processing the story relied on attention.

This analysis revealed the relative extent to which high-order regions have a stronger increase in response to stories when they are attended, compared to the facilitation of responses observed lower in the processing hierarchy in sensory areas (Fig. 10). Specifically, the response to text in the visual cortex and to speech in the auditory cortex was enhanced by attention (by up to 75%), but not as robustly as in cognitive regions (enhanced by 50% to 170%) such as the dPCC, the DLPFC, and the parahippocampal gyrus. Furthermore, calculating modulation level as a function of processing in a state of attention emphasizes subtle diversity in modulation strength between different high-level regions. For example, while attention significantly increased the processing of text in the right ventral prefrontal cortex (Fig. 9C vs. D), this increase was relatively small when considering the strength of text processing in the attended state in this region. The small ratio of modulation in this area is similar to the levels of modulations observed in high visual

cortices and is significantly lower than modulations in other high-order areas such as the dPCC (Fig. 10A).

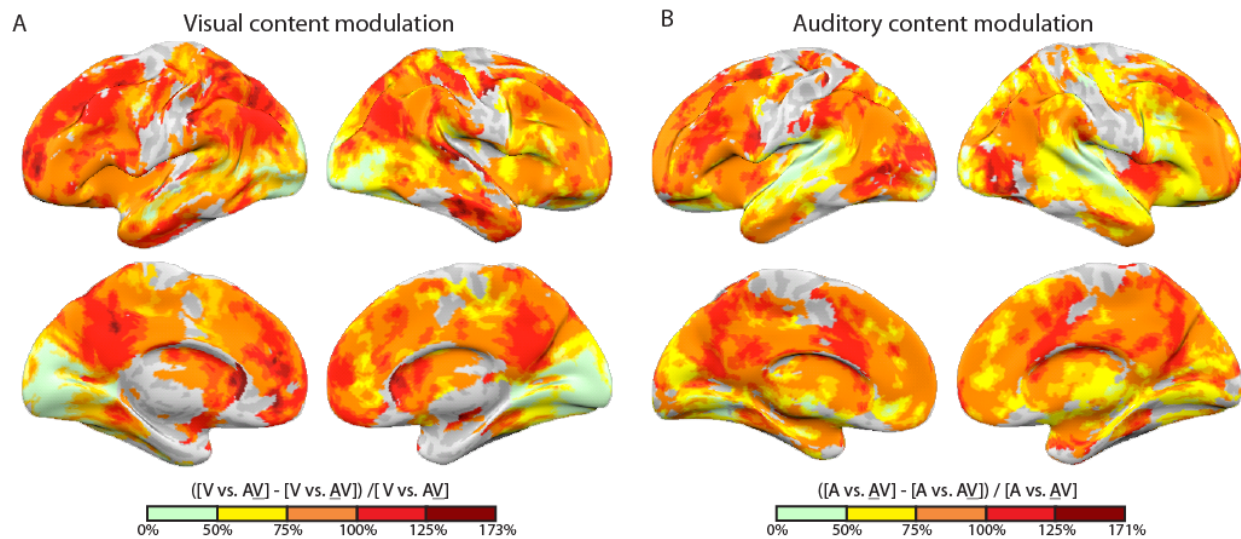


Figure 10. Brain regions exhibiting different degrees of attentional modulation of the neural response to spoken or written stories. **A**, Difference in ISC between attended and unattended states to written story as a percentage of the response to attended written story. **B**, Difference in ISC between attended and unattended states to spoken story as a percentage of the response to attended spoken story.

Regions of Interest

The results of the voxelwise analysis were reproduced in the group of 61 regions of interest.

High-order cognitive regions, such the left angular gyrus and the dPCC, responded to both the written and spoken stories, but only when they were attended (Fig. 9F). On the other hand, the auditory and visual regions responded to the stories in the corresponding modalities regardless of whether they were attended or not. However, despite their significance, the responses to the unattended stories were notably reduced compared to the states where the stories were attended (Fig. 9F).

Discussion

In this study, brain regions that process written and spoken real-life stories were mapped while either attended or ignored. Each local response was identified as containing task-relevant or irrelevant information by comparing it to its signature temporal response profile that is specific to the spoken or written story. Overall, this analysis revealed attention to have a similar hierarchical modulating effect on spoken and written content, in which auditory and visual sensory cortices processed unattended language, while high-order cognitive cortices did not (Fig. 9).

Despite the automatic response observed in early cortical regions, these regions were not completely indifferent to the state of attention. Responses to unattended spoken and written stories in the corresponding sensory cortices were weaker than when fully attended. The more limited bottom-up spread of irrelevant information into high-order regions could be attributed to the relatively diminished processing in sensory areas. This might represent different neural mechanisms that prevent more complex processing, such as a decrease in the population's SNR of the perceptual representation (Serences et al., 2009), or the lack of a specific processing component that is crucial for further downstream processing (e.g., low frequency oscillations, implicated in inter-regional communication; Stein and Sarnthein, 2000; Schroeder and Lakatos, 2009).

Nevertheless, some differences in the automatic processing of spoken and written stories were observed outside of the sensory cortices. A set of areas in the superior temporal cortex and inferior frontal gyrus responded to the spoken story even when unattended (Fig. 9B), but to the written story only when fully attended. These regions are known for their role in extracting semantic and lexical content (see Price, 2012) and were largely identified as having a modality-

invariant response profile in Chapter 2 (Fig. 5B). The involvement of these linguistic regions in processing unattended speech has been observed in a few previous studies (Jancke et al., 2001; Pulvermüller and Shtyrov, 2006; Rama et al., 2012).

The unattended written story was processed mostly in visual areas, including early medial occipital regions as well as higher lateral and ventral extrastriate regions (Fig. 9D). These results are consistent with previous studies that have shown automatic semantic processing of text in high-order visual regions (Dehaene et al., 2001; Peng et al., 2004; Kouider et al., 2007). Weaker traces of automatic processing of the written story were also observed in the posterior medial temporal cortex, which was also involved in processing unattended spoken content. This region, known for its role in accessing the semantic content of language (Price, 2012), has been observed in previous studies as part of an extended activation of the language network in unattended states (Price et al., 1996a; Peng et al., 2004).

The restricted processing of the unattended text compared to the unattended speech could be interpreted to suggest that written language requires more attention to achieve the same levels of linguistic processing as spoken language. This might reflect modality-specific neural mechanisms, such as the unique sensory pathways for reading or listening, or differences in cognitive control demands. However, this discrepancy in the automatic processing of spoken and written language might also result from small differences in measured eye movements across states of attention to text. While participants who ignored the text fixated their gaze on the written words, they did not seem to engage in short saccades within the words to the same degree as attentive readers. This might have slightly affected what visual information was available for further implicit processing in linguistic regions.

This study is limited in its ability to dissociate modality-specific effects from narrative-specific effects. Because the modality of the two stories was not counterbalanced, any differences observed between the neural processing of speech and text might be a result of their unique narratives rather than their sensory representation. Considering findings in Chapter 2 that showed an extended set of linguistic and other cognitive areas to have an invariant response to the modality of stories, effects in these areas (e.g., pSTG, precuneus, IFG) might be driven by the narrative. Nevertheless, neural effects in other frontal and parietal areas, as well as the sensory cortices, could be attributed to both the stories' unique linguistic content and the sensory modality.

The design of this study also doesn't allow stating whether the limited neural response to the unattended content was induced by the diversion of the focus of attention away from the irrelevant story or by the active attempt to ignore it. It could be hypothesized that the reason high-level areas did not process irrelevant information is their general modality-invariant role in the processing of stories (demonstrated in Chapter 2), which might not have the capacity for simultaneous cognitive operation on the two inputs. As such, the different traces of information in a crowded environment might compete to dominate these neural substrates in a winner-takes-all mechanism (Lavie et al., 2004). Yet since participants were explicitly instructed to ignore the task-irrelevant story, we cannot dismiss the possibility that engagement with one narrative does not influence the ability to process other irrelevant information presented simultaneously. In such a case, the limited processing of the irrelevant story observed here could have been a result of anticipatory suppression of distraction (Serences et al., 2004; Payne and Allen, 2011).

IV) Top-down attention modulation of inter-regional neural coupling of written and spoken language

As demonstrated in Chapter 3, even when written or spoken narratives are available in the environment, complete processing of complex linguistic content will not be induced automatically by the mere presence of the stimuli. Spoken and written content requires attention to be processed in high-order frontal and parietal areas. Without it, irrelevant language will be limited to early cortical processes. How do changes in the target of top-down attention prioritize task-relevant spoken and written input along the stable hierarchical neural pathways?

The encounter of the structural architecture of the brain with a given external environment has the potential to evoke different human behaviors. Given a specific context, dynamic cognition requires that the fixed structural architecture of the brain, namely its neural connections, support multiple functional architectures, or patterns of neural coupling. These functional architectures are the dynamic element that gives cognition flexibility and unlocks its potential to extract the relevant linguistic content. Investigating the dynamic unfolding over time of functional interactions is therefore key to understanding the brain as a substrate of the mind.

Top-down attention may help synchronize different processing levels, forming functional pathways that serve the processing of goal-relevant information (Miller and Cohen, 2001; Cohen et al., 2004). Such flexibility in information routing among cortical areas is especially important in a complex environment with multiple, sometimes contradicting traces of external input to be filtered for appropriate behavior. For example, modality-invariant cortical regions might not be able to process contradicting traces of information delivered simultaneously from multiple

sources. As a result, different top-down goals might be associated with changes in functional interactions.

Recent studies have looked at functional network dynamics by calculating the functional connectivity (FC) between cortical areas. In this method, the BOLD activity correlation is calculated between different regions, assessing their coupling during performance of tasks (e.g., spatial-visual/auditory, numerical computation, working memory) or rest. Patterns of FC showed differences across states of rest versus performance of various external tasks (Caclin and Fonlupt, 2006; Betti et al., 2013; Gonzalez-Castillo et al., 2015; Spadone et al., 2015). However, when experimental conditions vary from each other in both the stimulus presented as well as in the assigned cognitive task, it is difficult to assess the contribution of these two components to the changing functional patterns. Are these inter-regional dynamics mostly driven by the changing external input, or could they also be influenced by the current cognitive demands, independent of the stimulus?

In order to attribute the functional patterns to changes in states of attention, functional network dynamics should be tracked while keeping the external stimulus the same. By removing any external stimuli, a few studies were able to cleanly track changes in FC patterns due to differences in covert cognitive tasks as well as psychiatric conditions, measured at a state of rest (Craddock et al., 2009; Shirer et al., 2011; Krienen et al., 2014). Still, tracking FC dynamics attributed to top-down attention while a matching external stimulus such as language is presented serves a greater challenge: The similarity in the sensory input might have a strong effect, which might dominate the FC pattern over a more subtle effect of the internal cognitive state. Unsurprisingly, few have attempted this challenge, capturing some variance in FC patterns while

performing the same visual tracking task at different attention load levels (Alnæs et al., 2015; Kwon et al., 2017), but never while processing language.

To better characterize dynamics in functional interactions driven by changes in top-down attention within a stable but complex linguistic environment, a more refined approach is necessary. Inter-subject functional correlation (ISFC) is a powerful method for assessing inter-regional response similarity between different brains (instead of within a brain, as in FC; Simony et al., 2016). In calculating the inter-regional correlation between brains, the intrinsic neural dynamics (known to dominate FC patterns) are filtered out, as well as non-neural artifacts that can influence the correlation pattern within a brain. This procedure increases the signal-to-noise ratio for detecting inter-regional correlations induced by shared external stimuli across participants. While this analytical tool has been used in the past to accurately identify different functional patterns associated with exposure to scrambled vs. intact spoken narrative (Simony et al., 2016), its robustness has the potential to identify more nuanced dynamics in functional patterns resulting from changing internal goals alone.

In the current study, ISFC was harnessed to track dynamics in inter-regional integrations of spoken and written information relevant to the current goal, using the neural data set from Chapter 3 recorded from participants who were simultaneously presented with two unrelated stories, one spoken and one written. One group of participants was asked to attend to the written story while ignoring the spoken one, while the other group did the reverse. As before, their neural data were compared to two unimodal control groups who were exposed to only one of the stories and hence shared with the simultaneous groups the experience related to that particular story. Therefore, applying ISFC across brains from the simultaneous and control groups has the power to reveal inter-regional coupling driven by the specific shared story. As a result, in

addition to capturing general differences in functional patterns due to changing internal goals, this approach will also allow for the characterization of each functional interaction as goal-relevant or irrelevant.

Materials and Methods

Subject

The subject population was as described in Chapter 3.

MRI acquisition

Subjects were scanned in a 3T full-body scanner (Skyra, Siemens) as described in Chapter 3.

Stimuli and experimental design

To measure inter-regional couplings, brain activity of attending readers, attending listeners, unimodal readers and unimodal listeners was used, as in Chapter 3

To measure similarity in neural activity across the language modalities, brain activity of listeners and readers of a real-life story was used, as in Chapter 3.

Data analysis

fMRI data were preprocessed as in Chapter 3.

Intrinsic Connectivity Networks (ICNs) and Regions of Interest (ROI) analysis: We used the same 61 regions of interest defined in Chapter 3.

Inter-subject Functional Correlation (ISFC) analysis: We calculated the ISFC correlation matrix between all ROIs across brains (1) of subjects within the same condition (Simony 2015) and (2)

of subjects from two different conditions. The neural signals X_i measured from subject i , $i = 1, \dots, k$, are in the form of a $p \times n$ matrix that contains signals from p neural sources over n time points. All time courses were z-scored within subjects to zero mean and unit variance. Thus, the subject-based ISFC was calculated by the Pearson correlation between single subject and the average of all other subjects as:

$$C_i = \frac{1}{n} X_i \left[\frac{1}{k-1} \sum_{j \neq i} X_j^T \right]$$

Hence, the $p \times p$ group-based ISFC matrix was given by:

$$C = \frac{1}{k} \sum_i C_i$$

Subject-based ISFC between two conditions was calculated by the correlation between single subjects from one group and the average of all subjects from the other condition as:

$$\tilde{C}_i = \frac{1}{n} X_i \left[\frac{1}{k-1} \sum_i Y^T \right]$$

where the neural signals Y_i measured from subject i , $i = 1, \dots, k$. The cross-group ISFC matrix was given by:

$$\tilde{C} = \frac{1}{k} \sum_i \tilde{C}_i$$

The final ISFC matrix is given by $(C + C^T)/2$ within a group. This symmetry was imposed because the correlation between two brain regions was considered to be unidirectional, as in FC. Similarly, the final ISFC matrix between two groups is given $(\tilde{C} + \tilde{C}^T)/2$, and then averaged with the final ISFC matrix calculated between individuals from the second group and averaged signal from the first group.

The ISFC was performed across the following conditions: attention to speech condition vs. unimodal speech control condition, attention to text condition vs. unimodal text control condition, attention to text condition vs. unimodal speech control condition, attention to speech condition vs. unimodal text control condition, and unimodal text control condition vs. unimodal speech control condition. ISFC was also performed within each of the four groups.

To determine whether each measured inter-regional coupling was significant, it was compared to inter-subject correlations that were calculated between the unimodal readers and listeners at groups A and V. Because the participants in these groups were exposed to two different external stimuli with no common features (i.e., distinct modality and content), we assumed that the inter-regional correlation between these groups could be regarded as noise. Thus, inter-regional correlations were considered significant when they either exceeded the maximum ($\tilde{C} = 0.091$) or fell below the minimum ($\tilde{C} = -0.066$) correlation value measured between the two unimodal groups.

To minimize inter-regional correlations that might represent synchronization mediated by the external stimulus independent of anatomy, the measured ISFC values were filtered based on intrinsic neural architecture. For this purpose, FC analysis was performed between all ROIs, using data acquired in the resting state of an independent set of participants. Previous studies have demonstrated FC analysis of extended time courses to yield correlations that mostly represent the fixed structural architecture of the brain (Park and Friston, 2013). We chose to filter out any ISFC coupling between ROIs that did not also appear in the top one-third strongest FC correlations measured (i.e., FC correlations higher than 0.55).

Network visualization

For the data collected from the 61 regions, the 61×61 ISFC correlation matrix was represented in a circular graph with 61 nodes representing the ROIs. The thickness of lines between the nodes in the diagram indicates the strength of the correlation between them, while negative values are represented as dashed lines. To visualize the data, we used an in-house modification of the `circularGraph` function for MATLAB.

Results

Attentional modification of goal-related inter-regional coupling

Story-driven inter-regional coupling in states of attention

First, inter-regional functional coupling that is unique to the processing of the attended story was identified. To do so, the response time course of each unimodal control group was directly correlated to the response time course of the corresponding attention group (A subjects to AV subjects and V subjects to AV subjects). Because the unimodal control groups (A, V) were presented with only one story, their regional neural responses were treated as representing the “signature” temporal activation, characteristic to the (area-specific) processing of the presented story. Thus, by correlating the neural signature with the attention groups, story-specific inter-regional couplings that are driven by the processing of goal-relevant information could be identified. This inter-subject functional correlation (ISFC) between groups was computed within and between 61 independently defined regions of interest (Fig. 11, Table 3).

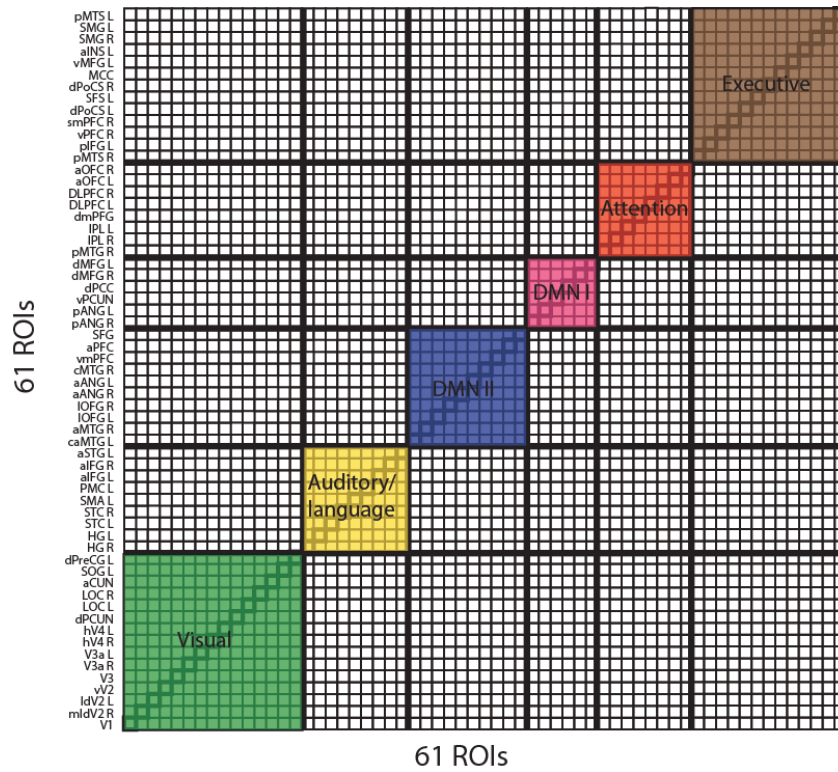


Figure 11. ISFC matrices structure. Correlations were calculated between the mean time courses of each of the 61 regions of interest, organized according to their ISC affiliation.

Statistical significance of ISFC was assessed by comparing the measured correlation to a noise distribution of inter-regional coupling between unimodal readers and listeners in groups A and V (Fig. 12A). Because participants in these two groups shared no common stimulus, we assumed that the inter-regional correlation between these groups could be regarded as noise. In addition, to minimize inter-regional correlations that might represent synchronization mediated by the external stimulus independent of anatomy, the measured ISFCs were filtered based on intrinsic neural architecture by measuring the functional connectivity between all ROIs during the resting state of an independent set of participants (see Materials and Methods). The cost of this conservative filtering step is a possible loss of some connections that are unique to focused attention (and not resting).

Inter-regional comparisons between groups of participants revealed extended interactions between regions sharing information from the attended stories. When attention was directed toward speech (AV), the entire cortical hierarchy was interconnected, revealing a pattern of inter-regional couplings that are specific to processing the spoken story (Fig. 12B, blue). Each of the explored regions of interest was correlated/anticorrelated with several other regions.

Although this coupling pattern represents shared responses that are typical to the processing of the speech, the interactions were not limited to a simple bottom-up processing hierarchy of auditory input, but also included all the visual ROIs.

When attention was directed toward text (AV), a similar (but far from identical) pattern of inter-regional couplings that are specific to processing the written story was observed (Fig. 12B, red).

While some interactions connected the same regions as when attended spoken content was shared, other interactions differed. For example, while most ROIs shared information related to the text with other regions, Heschl's gyri were not coupled to any other region or between them.

In contrast, these gyri shared information related to the attended speech with the STC and aIFGL. In addition, one must note that the inter-regional correlations seen across these two coupling patterns could not be treated as identical: the interaction between the same two regions could represent shared processing typical either to the spoken content or to the written one—depending on which story's signature response is correlated.

Story-driven inter-regional coupling in states of deviated attention

Next, the same analysis was repeated, but response time courses were correlated from each unimodal group to the opposite attentional group (V correlated to AV; A correlated to AV).

These comparisons were designed to reveal inter-regional couplings that are driven by the processing of the goal-irrelevant information (from the unattended stories).

Inter-regional comparisons between groups of participants revealed limited interactions between low-level regions, sharing information from the unattended stories. When internal goals differed and the spoken story was ignored (AV), the coupling pattern representing shared spoken content was substantially different. In this cognitive state, the regions that maintain speech-specific interactions were confined to the auditory cortices and some linguistic areas (Fig. 12C, blue). These regions included Heschl's gyri, STC, PMC L, aSTG L, caMTG L, aMTG R, cMTG R, aIFG, and IOFG. The significant inter-regional correlations measured here were positive and varied in their strength. When the written story was ignored (AV), the coupling pattern representing shared written content was also greatly confined, with remaining significant interactions limited to the medial visual cortices (Fig. 12C, red). These regions included V1, mldV2 R, vV2, V3a L, and aCUN.

Together, these results demonstrate the influence of internal goals on inter-regional interaction dynamics. Directing attention toward a story was associated with the widespread dominance of functional interactions specific to that attended (spoken or written) story. At the same time, information from the simultaneously presented but unattended story was also shared, but only between regions in the corresponding sensory cortices.

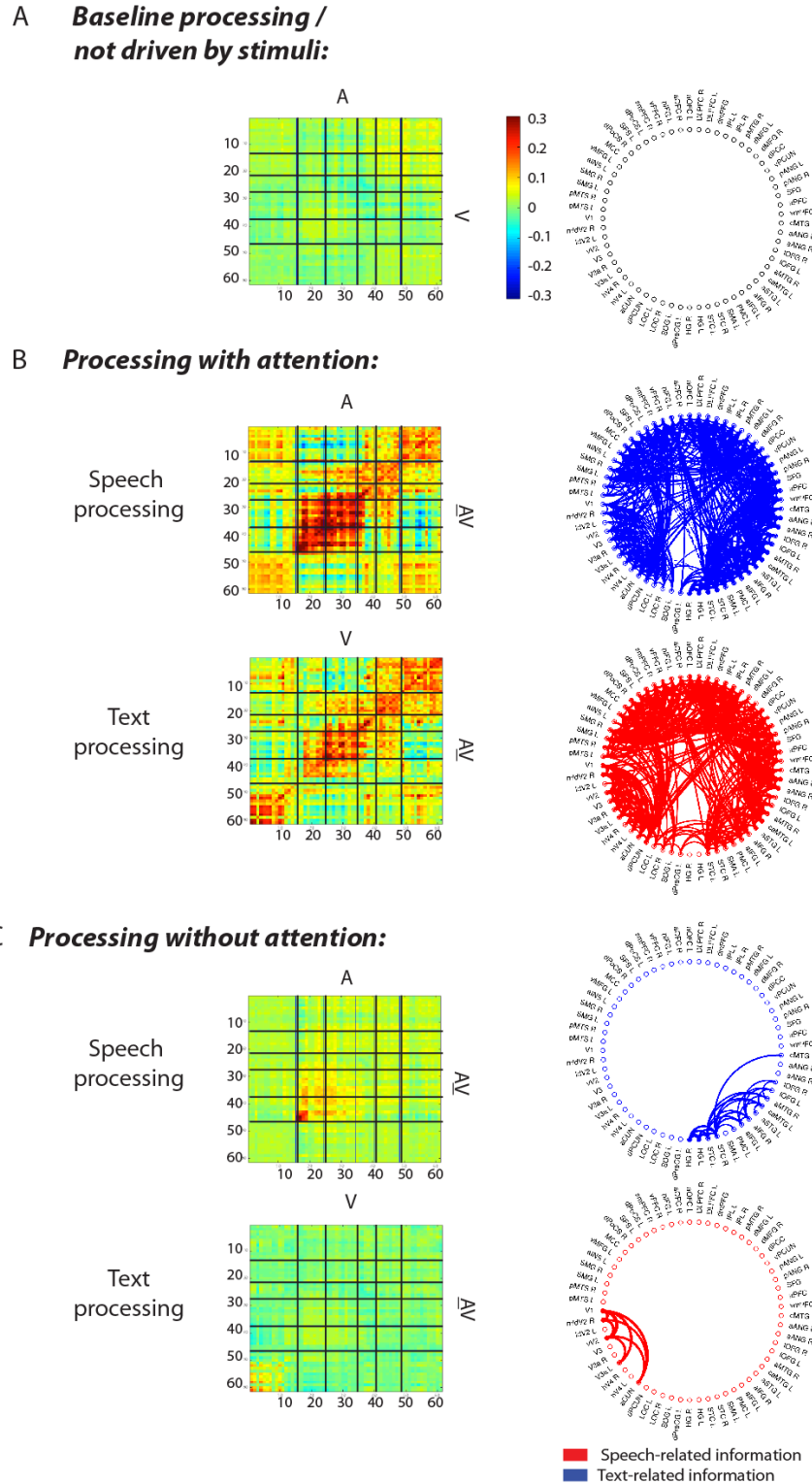


Figure 12. ROI-based ISFC correlation matrices reveal top-down modulation by attention of story-specific inter-regional interactions. **A**, The ISFC correlation matrix was empty when comparing unimodal listeners (A) and readers (V). **B**, ISFC correlation matrix between unimodal listeners (A) and multisensory group who attended speech (AV), and between unimodal readers (V) and multisensory group who attended text (AV). **C**, ISFC correlation matrix between unimodal listeners (A) and

multisensory group who did not attend speech (AV), and between unimodal readers (V) and multisensory group who did not attend text (AV). The circle connectivity presents threshold ISFC levels between the ROIs. Line width represents the log of the pairwise correlation between different regions.

Discussion

In this study, we were able to track the dynamics of inter-regional coupling patterns as a function of current internal goals. The direction of top-down attention toward or away from stories modulated which brain areas interacted and whether those interactions were driven by shared response to the attended or unattended story. Specifically, attention influenced the extent to which story-related content was shared along the processing hierarchy: although story-related responses were shared between areas in sensory cortices even when outside of the focus of attention, top-down attention was required for the story-specific response to be shared also with higher, more cognitive frontal and parietal areas.

The decreased functional coupling with high cortical areas in response to the unattended story was not due to a complete lack of cortical processing of that content. Chapter 3 showed that unlike frontal and parietal areas, lower sensory cortices and some linguistic regions had a neural response that is typical for the processing of the stories, regardless of whether the story was attended or not. For example, V2 and STC had a significant response to the written and spoken stories correspondingly in both states of attention, but were coupled with frontal and parietal regions only when the modality-appropriate story was attended.

Previous methodological approaches have successfully been used to study the effect of changes in top-down attention on inter-regional interactions, mainly within the visual processing hierarchy. Several studies have regressed out stimulus-evoked responses and examined

“background connectivity” in the residuals, which improved the sensitivity to intrinsic interactions between cortical regions of interest (Al-Aidroos et al., 2012; Norman-Haignere et al., 2012; Griffis et al., 2015; Cordova et al., 2016). To determine whether a certain neural coupling is goal-relevant, these studies have usually categorized the observed coupling based on previously acquired knowledge about the interacting regions of interest. For example, increased background connectivity between the fusiform face areas (FFA) and V4 was categorized (Al-Aidroos et al., 2012) as an increased goal-relevant response when faces were attended, based on the known role of FFA in face processing (Kanwisher, 2010).

Here, it has been demonstrated for the first time that ISFC could serve as a complementary method to study the effects of changes in top-down attention on inter-regional coupling. By filtering out spontaneous neural responses, which contribute strongly to FC, the sensitivity to stimulus-locked processes is improved (Simoney et al., 2016), enabling the identification of distinct functional patterns in different states of attention. This effect of attention on functional patterns of coupling was previously attenuated when using FC (Alnæs et al., 2015; Kwon et al., 2017). In addition, by following inter-regional couplings characteristic of each story, this study was uniquely able to show connectivity that is driven specifically by goal-relevant information across the entire processing hierarchy. This approach is especially useful for tracking changes in coupling between areas without a well understood functional role, for which modeling the information that should be considered as goal-relevant can be challenging (e.g., in contrast to early visual cortical areas).

General Summary

In this dissertation, novel analytical tools have been used to explore commonalities and divergences of neural processes across the spoken and written language modalities. First, it was observed that real-life written and spoken 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 (Fig. 3). Further on, modality-invariant (Fig. 5) and modality-selective (Fig. 4) responses to written and spoken narratives were identified by directly comparing response time courses across listeners and readers.

By comparing the responses across modalities, we observed a remarkable invariance to the form of input (written and spoken narratives) outside of the sensory cortices (Fig. 5). Such modality-invariant responses were measured in linguistic areas in the pSTG and the IFG bilaterally, as well as in extra-linguistic cognitive regions such as the precuneus and anterior parts of the dmPFC. Prior studies have 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., 2002b; Marinković, 2004; Spitsyna et al., 2006; Jobard et al., 2007; Lindenberg and Scheef, 2007; Vagharchakian et al., 2012). Moreover, the shared responses across spoken and written language extended in this experiment 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 that their representations are highly abstracted from the sensory input.

The similarity of response time courses to spoken and written narratives attests to the plasticity of the language system. Linguistic regions that exhibited 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; Calver 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.

The strong modality-invariance in these brain areas was accompanied by modality-selective responses in sensory cortices as well as high-order parietal and frontal cortices (Fig. 4). In particular, a greater reliability was observed 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 the anterior ventrolateral prefrontal cortex (VLPFC) and for vocalizations in the 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 the medial and lateral frontal cortex exhibited a preference for spoken over written language (and vice versa) 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., auditory 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 responded reliably to the meaningful linguistic stimuli and not to unintelligible scrambled letters or sounds. These findings might 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.

Next, brain regions that process written and spoken real-life stories and their inter-regional couplings were mapped in different states of attention to language. Using ISC analysis, each local response was identified as containing task-relevant or irrelevant information by comparing it to its signature temporal response profile that is specific to the spoken or written story.

Similarly, using ISFC, we tracked task-relevant or irrelevant inter-regional dynamics by

comparing each region's temporal response to the signature response specific to the spoken or written story in other regions of the brain. These analytical approaches revealed attention to have a similar hierarchical modulation effect on spoken and written content, in which auditory and visual sensory cortices processed unattended language, unlike high-order cognitive cortices (Figs. 9,10). But although story-related responses were shared between areas in sensory cortices even when outside the focus of attention, top-down attention was required for the story-specific response to be shared also with higher, more cognitive frontal and parietal areas (Fig. 12).

Despite the automatic response observed in early cortical regions, they were not completely indifferent to the state of attention. Responses to unattended spoken and written stories in the corresponding early cortices were weaker than when fully attended. The more limited bottom-up spread of irrelevant information into high-order regions could be attributed to the relatively diminished processing in sensory cortices. This might represent different neural mechanisms that prevent more complex processing of the irrelevant input, such as a decrease in the population's SNR of the perceptual representation (Serences et al., 2009), or the lack of a specific processing component that is crucial for further downstream processing (e.g., low frequency oscillations, which have been implicated in inter-regional communication; Stein and Sarnthein, 2000; Schroeder and Lakatos, 2009). Nevertheless, while the decreased response in sensory cortices might be the reason for the decreased functional coupling with high-order areas, it did not prevent some of the inter-regional interactions within sensory areas. For example, V2 and STC had a significant response to the written and spoken stories correspondingly in both states of attention (Fig. 9F), and were connected to other sensory cortices in both states of attention (Fig. 12). However, they were coupled with frontal and parietal regions only when the modality-appropriate story was attended (Fig. 12C).

Some differences in the automatic processing of spoken and written stories were observed outside the sensory cortices. A set of areas in the STC and IFG responded to the spoken story even when unattended (Fig. 9B), but to the written story only when fully attended (Fig. 9C). The involvement of these linguistic regions in processing unattended speech was observed in a few previous studies (Jancke et al., 2001; Pulvermüller and Shtyrov, 2006; Rama et al., 2012). The unattended written story was processed mostly in visual areas, including early medial occipital regions as well as higher lateral and ventral extrastriate regions (Fig. 9D). These results are consistent with previous studies that have shown an automatic semantic processing of text in high-order visual regions (Dehaene et al., 2001; Peng et al., 2004; Kouider et al., 2007). Weaker traces of automatic processing of the written story were also observed in the posterior MTG, which was also involved in processing of unattended spoken content. This region, known for its role in accessing the semantic content of language (Price, 2012), has been reported in previous studies as part of an extended activation of the language network in unattended states (Price et al., 1996a; Peng et al., 2004). The restricted processing of the unattended text compared to the unattended speech could be interpreted as suggesting written language requires more attention to achieve the same levels of linguistic processing as spoken language. This might reflect modality-specific neural mechanisms, such as the unique sensory pathways for reading vs. listening, or differences in cognitive control demands.

General Discussion

Taken together, the results of this study suggest that the human brain relies heavily on the neural mechanisms of spoken language in order to accommodate the skill of literacy. As in spoken word comprehension, written language starts cortical processing within its corresponding sensory cortex, where the unique sensory features of the linguistic input are processed automatically and shared locally without requiring focused attention. Further downstream, outside sensory real estate, the top-down effect of attention on processing of both spoken and written content increases and peaks at frontal and parietal regions (e.g., PCC and DLPFC). Hence, it appears that most high-order regions do not process spoken or written language automatically and share story-specific content with the rest of the brain only when it is fully attended. Some of these high-order regions with attention-dependent response to stories also showed modality-invariant responses to stories (e.g., PCC, anterior dmPFC, angular gyrus).

These neural similarities in comprehension of spoken and written language could imply that a successful acquisition of reading skills involves the appropriation of some neural processes, previously thought to be characteristic of spoken language alone. The repurposing of neural mechanisms underlying spoken language is not surprising when considering that writing systems have been in wide use only as of the last century, which makes the existence of naturally adapted brain mechanisms that specialize in written language unlikely (Gleitman and Rozin, 1977; Liberman, 1992). Regardless of whether evolution or development influenced the functional structure of the brain to allow reading, according to neural reuse theories, it is quite common for neural circuits established for one purpose to acquire new uses, which do not need to involve much local change to the original circuit structure (e.g., it might involve only the establishment of functional connections to new neural partners) (Anderson, 2010).

Moreover, speech in its natural form has always been highly intertwined with visual information from the speaker's face, which can establish a neural scaffold for cross-modal interactions in language processing. Behaviorally, lip movements have been shown to improve speech perception, especially in a noisy environment (Sumby and Pollack, 1954; McGurk and MacDonald, 1976; Campanella and Belin, 2007). Many neuroimaging studies have shown that visual information from the speaker can by itself activate the auditory cortex (Molholm and Foxe, 2005; Campbell, 2008), or modulate speech processing during stages previously assumed to be unimodal auditory (Calvert et al., 1999; Amedi et al., 2005). Thus, a network that already has the neural structure necessary for audio-visual interaction of language-related information might serve as a “neuronal niche”, supporting the novel culturally acquired cognitive ability (Dehaene, 2005; Dehaene and Cohen, 2007).

Further evidence for the dependency of literacy on cognitive components of speech comprehension could also be found in behavioral findings related to reading acquisition. Reading proficiency typically follows the acquisition of spoken language, and the success of this learning process has been shown to depend on phonemic skills because their association of graphemes is crucial for successful reading (Gough and Hillinger, 1980; National Reading Panel, 2000; Perfetti and Sandak, 2000; Snowling and Hulme, 2005). Deficiencies in reading could later be treated through the practice of phonologic awareness, as has been shown in healthy and dyslexic children (Lovett et al., 1994; Ehri et al., 2001). Furthermore, the tight link of reading to auditory ability could be seen in the reported low literacy levels of deaf people (Perfetti and Sandak, 2000, but see Mayberry et al., 2010).

Altogether, some of our neural findings, as well as previous neural evidence and behavioral reports, suggest reading to be to a large extent “parasitic to speech” (Mattingly, 1972).

Nonetheless, the neural similarities observed here between comprehension of spoken and written language might not be a result of the processes of literacy acquisition. Speech is not the only auditory information that could be influenced by the visual modality—visual influences in the “auditory” cortex have been observed across species and for many types of stimuli (Schroeder and Foxe, 2005; Kayser et al., 2007). A shared neural response across modalities could hypothetically occur when comparing temporally aligned stimuli that share a semantic meaning (e.g., a visual object and its auditory manifestation). Future exploration of the temporal profiles of neural responses to written and spoken language before and during acquisition of reading skills might pinpoint cross-modal similarities that result from the plasticity in the language network. Characterizing the development of modality-invariant neural response may allow us to better understand and treat reading deficits, as well as point toward more effective pedagogical strategies to achieve literacy.

At the same time, in addition to the commonalities in the processing of speech and text, the results of this study also point toward some neural distinctions and modality-specific effects. Although some linguistic areas (i.e. pSTG, angular gyrus, IFG) were found to have a modality-invariant response, only information from spoken stories was processed automatically in these regions and shared between them when outside the focus of attention. Unattended written language was processed and shared mainly within occipital sensory regions, reaching only the pMTG. In addition, functional selectivity was observed in some frontal and parietal areas that may indicate modality-specific cognitive control processes.

These neural disparities in comprehension of spoken and written language could imply that a successful acquisition of reading skills might require not just common neural mechanisms, but also some neural processes that differ from the ones involved in spoken language. This should

not be surprising, given the behavioral differences in the learning processes of spoken vs. written language. In contrast to acquiring speech, achieving literacy requires active, effortful learning and depends on the availability and adequate development of some cognitive control processes (such as attention and working memory; Daneman and Blennerhassett, 1984). Reading difficulties are also much more prominent, specifically in people with attentional and memory deficits (e.g., in ADD and ADHD; American Psychiatric Association, 1994; Felton et al., 1987; Dykman and Ackerman, 1991; Fergusson and Horwood, 1992; Purvis and Tannock, 1997). These evidences might hint at different cognitive control functions being involved in reading vs. listening.

Previous studies have painted a picture of a more complex relationship between spoken and written language, with some neural components possibly unique to literacy or speech comprehension. Only the written modality is thought to have a sensory word form area (VWFA; Cohen et al., 2000; Dehaene et al., 2001), unlike spoken language (Wise et al., 2001; Narain et al., 2003; Specht and Reul, 2003; Thierry et al., 2003; Benson et al., 2006). The sensory differences between the modalities also enhance the association of some cognitive processes more strongly with reading than listening, and could be reflected in neural activation in some cortical areas. For example, in successful reading from a page (but not in listening), the eye must execute a series of accurate saccadic movements along the text (Rayner, 1978), which are manifested in neural activation in the FEF (Michael et al., 2001; Choi et al., 2014). We were not able to see such modality-specific response in our results due to the rapid serial visual presentation of the text required by our analytical methods. Moreover, spatial direction of attention toward the written input is essential for reading (Brannan and Williams, 1987; Facoetti et al., 2000; 2001), and therefore could evoke activation in posterior parietal regions involved in

spatial attention (Brunn and Farah, 1991; Hari and Renvall, 2001). Although these processes are not unique to language comprehension and could be found in many human behaviors, they are strongly associated with the task of reading and less so with comprehension of spoken language.

Speech comprehension might also have its own characteristic cognitive processes. Spoken messages rely on several extra-linguistic factors to determine the total significance of the message, such as voice intonation and stress, gesture, shared contextual knowledge, and to some degree pragmatic implicature (Grimes, 1972). For example, previous studies have demonstrated processing of intonational information during speech in the right superior temporal sulcus, dorsolateral, and orbitofrontal areas (mirroring the traditionally left lateralized language network) (e.g. Buchanan et al., 2000; Kotz et al., 2003; Wildgruber et al., 2005). One of the essential functions of prosody is to provide a basic cognitive skeleton that allows one to hold an auditory sequence in working memory until a more complete semantic analysis can be carried out (Koriat et al., 2002; Frazier et al., 2006; Swets et al., 2007). In contrast to oral messages, written statements do not contain this additional information and must be explicit and context free, because the reader is removed from the writer by time and space. Interestingly, although our results did not show a prosody-driven biased response to speech in the right hemisphere, a speech selective response was observed in frontal and parietal areas that could reflect speech-specific memory and/or attention networks (Protzner and McIntosh, 2007; Braga et al., 2013). The lack of right lateralization in the language network could be explained by possible “implicit prosody” imagined in silent reading (e.g. Ashby, 2006; Hirotani et al., 2006), which was possibly manifested in the same location and with the same temporal response profile as vocalized prosody.

The sensory and cognitive components unique to the task of reading or listening are difficult to

disentangle from each other. For example, complex emotions could not be conveyed through the visual features of text, while a voluntary movement of the ears could not be used in speech comprehension. Therefore, the word “modality” in this dissertation subsumes both sensory modality (auditory vs. visual) and task modality (listening vs. reading). Furthermore, the effects depicted as specific to the written or spoken stories in Chapters 3 and 4 could not be inferred to be driven only by the language modality. This is a limitation of the experiment design, which did not counterbalance the language modality and the narratives, and therefore the effects of these two factors could not be dissociated.

The act of reading has been difficult to measure in neuroscience due to methodological issues and reading pace variability across participants. Traditional neuroimaging techniques allow us to observe the structure and function of the nervous system, but such research typically employs simplified versions of human linguistic experiences, for example, imaging the brain of a person viewing random words or sentences. These highly controlled experimental setups often fall short of explaining real-life human cognition, where the components of our environment constantly interact, creating a complex multi-layered experience. Aggregated responses to such isolated language can fail to map the full set of regions engaged in real-life language comprehension and ongoing attention in a complex environment, specifically in high-order regions (Lerner et al., 2011; Ben-Yakov et al., 2012). In this dissertation, a more naturalistic approach was attempted by presenting participants with written and spoken versions of engaging stories, which allowed for the capture of neural activity underlying spoken and written language across the entire processing hierarchy in its full real-world complexity.

ISC analysis (Hasson et al., 2009) provides an ideal tool for measuring the similarity of the response time course to complex spoken and written stimuli. The demonstration of spatially

overlapping neural responses is not strong evidence for modality or attention-level invariance. In principle, a brain area could respond reliably to both spoken and written narratives or both states of attention, but with one temporal response profile for the spoken narrative (or attended state) and another for the written narrative (or unattended state). This dissertation goes beyond previous results by demonstrating that the response time courses for real-life complex narratives across the two conditions of modality and two attention states were highly similar within the regions noted above.

Although our ISC analysis method is successful at characterizing the neural dynamics that are shared over time across spoken and written or attended and unattended 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).

Tracking how top-down attention prioritizes task-relevant spoken and written input along the fixed hierarchical neural structure required adopting a novel methodological approach. Inter-regional functional interactions are commonly investigated using an analysis termed “functional connectivity” (FC), by computing the temporal covariance of the BOLD signal across distinct brain regions. Because this is performed within a single subject’s brain, each covariance value consists of the aggregate of three sources of relationships: First, intrinsic neural interactions, defined as responses arising from the anatomical relationship between the two regions, or from intrinsic cognitive process unrelated to ongoing stimulus processing; second, neural covariance

driven by the shared synchronization of the two different regions to the same external stimulus presented; last, non-neural artificial covariance, like respiration rate or heartbeat. The strong contribution of intrinsic neural and non-neural interactions has shown FC patterns to be very similar when different stimuli are presented to participants (Simony et al., 2016). Therefore, FC analysis is not always sensitive enough to capture dynamic stimulus-induced variance, let alone smaller changes in external stimulus processing in a complex but stable environment.

ISFC analysis is a powerful method for assessing inter-regional response similarity between different brains (instead of within a brain, as in FC; Simony et al., 2016). In calculating the inter-regional correlation between brains, the intrinsic neural dynamics are filtered out, as well as non-neural artifacts that can influence the correlation pattern within a brain. This procedure increases the signal-to-noise ratio in detecting inter-regional correlations induced by shared external stimuli across participants. While this analytical tool has been used in the past to clearly identify different functional patterns associated with exposure to various external stimuli (Simony et al., 2016), its robustness allowed this study to identify more nuanced dynamics in functional patterns resulting from changing internal goals alone. Thus, this study has been able to show for the first time the role of large-scale functional visual and auditory neural pathways in active language processing, and their interactions.

When applying a correlation-based methodology to study inter-regional functional interactions, the question of causality must be carefully considered before drawing conclusions. High correlation between remote sites in the brain might imply some direct neural connection, a third site driving their joint activation, or just a correlation of the regions with an external stimulus. Even when there is an observed correlation between two sites, it is extremely difficult to determine the direction of the connection (i.e., the direction of the flow of information). While

the current study successfully identified regions participating in processing the spoken and written narratives in different states of attention, it is not able to determine whether one region directly sends information to another. Therefore, while these findings uncover the groundwork of the functional network architecture underlying auditory and visual linguistic comprehension—and open the door to informed hypotheses about the direction of causality between brain regions—definite causal conclusions must be tested in subsequent research using techniques such as transcranial magnetic stimulation or human brain lesion cases.

In this dissertation, new analytical tools have been employed to explore how the human brain accommodates our newly developed skill of literacy with our evolutionarily older ability to comprehend speech. Comparison of the temporally extended neural response across spoken and written real-life stories, in attended and unattended states, allowed for the mapping of commonalities and divergences of neural processes across the language modalities. Overall, the results presented in this dissertation argue that our ability to extract the same information from spoken and written forms arises from a mixture of both selective and shared neural processes. This attests to the dynamic nature of the brain, which acquires the ability to use the same neural substrate to process different inputs according to current internal demands while adjusting other neural functions to serve the unique demands of the specific task.

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

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. 3A,B); modality-selective responses to the spoken condition (S > W) and to the written condition (W > S; Fig. 4A, red and green); inter-SC across the spoken and written conditions (S + W; Fig. 5B). 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.

Table 3 - 61 ROIs

Network	Area	Abbreviation
Visual	V1	V1
	dorsal V2 medial and lateral (right)	mldV2 R
	dorsal V2 lateral (left)	ldV2 L
	ventral V2	vV2
	V3	V3
	V3a (right)	V3a R
	V3a (left)	V3a L
	hV4 (right)	hV4 R
	hV4 (left)	hV4 L
	anterior cuneus	aCUN
	dorsal precuneus	dPCUN
	lateral occipital cortex (left)	LOC L
	lateral occipital cortex (right)	LOC R
	superior occipital gyrus (left)	SOG L
	dorsal precentral gyrus (left)	dPreCG L
Auditory/language	Heschl's gyrus (right)	HG R
	Heschl's gyrus (left)	HG L
	superior temporal cortex (left)	STC L
	superior temporal cortex (right)	STC R
	supplementary motor area (left)	SMA L
	premotor cortex (left)	PMC L
	anterior inferior frontal gyrus (left)	aIFG L
	anterior inferior frontal gyrus (right)	aIFG R
DMN II	anterior superior temporal gyrus (left)	aSTG L
	central-anterior middle temporal gyrus (left)	caMTG L
	anterior middle temporal gyrus (right)	aMTG R
	lateral orbitofrontal gyrus (left)	IOFG L
	lateral orbitofrontal gyrus (right)	IOFG R
	anterior angular gyrus (right)	aANG R
	anterior angular gyrus (left)	aANG L
	central middle temporal gyrus (right)	cMTG R
	ventral medial prefrontal cortex	vmPFC
	anterior prefrontal cortex	aPFC
	superior frontal gyrus	SFG
DMN I	posterior angular gyrus (right)	pANG R
	posterior angular gyrus (left)	pANG L
	ventral precuneus	vPCUN
	dorsal posterior cingulate cortex	dPCC
	dorsal middle frontal gyrus (right)	dMFG R

	dorsal middle frontal gyrus (left)	dMFG L
Attention	posterior middle temporal gyrus (right)	pMTG R
	inferior parietal lobule (right)	IPL R
	inferior parietal lobule (left)	IPL L
	dorsal medial prefrontal gyrus	dmPFG
	dorsolateral prefrontal cortex (left)	DLPFC L
	dorsolateral prefrontal cortex (right)	DLPFC R
	anterior orbitofrontal cortex (left)	aOFC L
	anterior orbitofrontal cortex (right)	aOFC R
Executive	posterior inferior frontal gyrus (left)	pIFG L
	ventral prefrontal cortex (right)	vPFC R
	superior medial prefrontal cortex	smPFC
	dorsal postcentral sulcus (left)	dPoCS L
	superior frontal sulcus (left)	SFS L
	dorsal postcentral sulcus (right)	dPoCS R
	mid-cingulate cortex	MCC
	ventral middle frontal gyrus (left)	vMFG L
	anterior insula (left)	aINS L
	supramarginal gyrus (right)	SMG R
	supramarginal gyrus (left)	SMG L
	posterior middle temporal sulcus (right)	pMTS R
	posterior middle temporal sulcus (left)	pMTS L

References

- Al-Aidroos N, Said CP, Turk-Browne NB (2012) Top-down attention switches coupling between low- level and high-level areas of human visual cortex. *PNAS* 109:14675–14680.
- Alnæs D, Kaufmann T, Richard G, Duff EP, Sneve MH, Endestad T, Nordvik JE, Andreassen OA, Smith SM, Westlye LT (2015) Attentional load modulates large-scale functional brain connectivity beyond the core attention networks. *NeuroImage* 109:260–272.
- Amedi A, Kriegstein von K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research* 166:559–571.
- American Psychiatric Association (1994) Diagnostic Criteria from DSM-IV, Washington, DC: American Psychiatric Association.
- Anderson ML (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Science* 33:245–313.
- Ashby J (2006) Prosody in skilled silent reading: evidence from eye movements. *J Research in Reading* 29:318–333.
- Baldassano C, Beck DM, Fei-Fei L (2015) Parcellating connectivity in spatial maps. *PeerJ* 3:e784.
- Banich MT, Milham MP, Atchley R, Cohen NJ, Webb A, Wszalek T, Kramer AF, Liang Z-P, Wright A, Shenker J, Magin R (2000) fMRI Studies of Stroop Tasks Reveal Unique Roles of Anterior and Posterior Brain Systems in Attentional Selection. *Journal of Cognitive Neuroscience* 12:988–1000.
- Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. *Cognitive Brain Research* 13:17–26.
- Ben-Yakov A, Honey CJ, Lerner Y, Hasson U (2012) Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage* 63:1–6.
- Benson RR, Richardson M, Whalen DH, Lai S (2006) Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *NeuroImage* 31:342–353.
- Besner D, Stolz JA (1999) Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin & Review* 6:449–455.
- Bestelmeyer PEG, Belin P, Grosbras M-H (2011) Right temporal TMS impairs voice detection. *CURBIO* 21:R838–R839.
- Betti V, Penna Della S, de Pasquale F, Mantini D, Marzetti L, Romani GL, Corbetta M (2013) Natural Scenes Viewing Alters the Dynamics of Functional Connectivity in the Human Brain. *Neuron* 79:782–797.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience* 7:295–301.
- Bluvas EC, Gentner TQ (2013) Attention to natural auditory signals. *Hearing Research* 305:10–18.

- Bolger DJ, Hornickel J, Cone NE, Burman DD, Booth JR (2008a) Neural correlates of orthographic and phonological consistency effects in children. *Human Brain Mapping* 29:1416–1429.
- Bolger DJ, Minas J, Burman DD, Booth JR (2008b) Differential effects of orthographic and phonological consistency in cortex for children with and without reading impairment. *Neuropsychologia* 46:3210–3224.
- Booth JR, Burman DD, Meyer JR, Gitelman DR (2002a) Functional Anatomy of Intra- and Cross-Modal Lexical Tasks. *NeuroImage*.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002b) Modality independence of word comprehension. *Human Brain Mapping* 16:251–261.
- Booth JR, Lu D, Burman DD, Chou T-L, Jin Z, Peng D-L, Zhang L, Ding G-S, Deng Y, Liu L (2006) Specialization of phonological and semantic processing in Chinese word reading. *Brain Research* 1071:197–207.
- Braga RM, Wilson LR, Sharp DJ, Wise RJS, Leech R (2013) Separable networks for top-down attention to auditory non-spatial and visuospatial modalities. *NeuroImage*:1–10.
- Brainard DH (1997) The Psychophysics Toolbox. *Spatial Vision* 10:433–436.
- Brannan JR, Williams MC (1987) Allocation of visual attention in good and poor readers. *Perception & Psychophysics* 41:23–28.
- Brem S, Bach S, Kucian K, Kujala JV, Guttorm TK, Martin E, Lyytinen H, Brandeis D, Richardson U (2010) Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *PNAS* 107:7939–7944.
- Brunn JL, Farah MJ (1991) The Relation Between Spatial Attention and Reading: Evidence from the Neglect Syndrome. *Cognitive Neuropsychology* 8:59–75.
- Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research* 9:227–238.
- Caclin A, Fonlupt P (2006) Functional and effective connectivity in an fMRI study of an auditory-related task. *European Journal of Neuroscience* 23:2531–2537.
- Calvert GA (2001) Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. *Cerebral Cortex* 11:1110–1123.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport* 10:2619–2623.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *CURBIO* 10:649–657.
- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends in Cognitive Sciences* 11:535–543.
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:1001–1010.

- Castelhano MS, Muter P (2001) Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology* 20:237–247.
- Catena A, Fuentes LJ, Tudela P (2002) Priming and interference effects can be dissociated in the Stroop task: New evidence in favor of the automaticity of word recognition. *Psychonomic Bulletin & Review* 9:113–118.
- Chawla D, Rees G, Friston KJ (1999) The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience* 2:671–676.
- Chee MWL, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and Visual Word Processing Studied With fMRI. *Human Brain Mapping* 7:7–15.
- Chee MWL, Weekes B, Lee KM, Soon CS, Schreiber A, Hoon JJ, Chee M (2000) Overlap and Dissociation of Semantic Processing of Chinese Characters, English Words, and Pictures: Evidence from fMRI. *NeuroImage* 12:392–403.
- Choi W, Desai RH, Henderson JM (2014) The neural substrates of natural reading: a comparison of normal and nonword text using eyetracking and fMRI. *Frontiers in Human Neuroscience* 8:22.
- Choi W, Henderson JM (2015) Neural correlates of active vision: An fMRI comparison of natural reading and scene viewing. *Neuropsychologia* 75:109–118.
- Cohen J, Aston-Jones G, Gilzenrat MS (2004). A Systems-Level Perspective on Attention and Cognitive Control: Guided activation, adaptive gating, conflict monitoring, and exploitation vs. exploration. In: Posner MI (Ed.), *Cognitive Neuroscience of Attention* (pp 71–90). New York: Guilford Press.
- Cohen L, Dehaene S, Naccache L, Lehericy S, Dehaene-Lambertz G, Henaff M-A, Michel F (2000) The visual word form area. *Brain* 123:291–307.
- Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* 125:1054–1069.
- Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and Divided Attention during Visual Discriminations of Shape, Color, and Speed: Functional Anatomy by Positron Emission Tomography. *The Journal of Neuroscience* 11:2383–2402.
- Cordova NI, Tompary A, Turk-Browne NB (2016) Neurobiology of Learning and Memory. *Neurobiology of Learning and Memory* 134:115–122.
- Craddock RC, Holtzheimer PE III, Hu XP, Mayberg HS (2009) Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine* 62:1619–1628.
- Crozier S, Sirigu A, Lehericy S, van de Moortele P-F, Pillon B, Grafman J, Agid Y, Dubois B, LeBihan D (1999) Distinct prefrontal activations in processing sequence at the sentence and script level: An fMRI study. *Neuropsychologia* 37:1469–1476.
- Daneman M, Blennerhassett A (1984) How to Assess the Listening Comprehension Skills of Prereaders. *Journal of Educational Psychology* 76:1372–1381.
- Dehaene S, Le Clec'H G, Poline J-B, Le Bihan D, Cohen L (2002) The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *NeuroReport* 13:321–325.

- Dehaene S (2005) Evolution of human cortical circuits for reading and arithmetic: The “neuronal recycling” hypothesis. In: From monkey brain to human brain, ed. Dehaene S, Duhamel JR, Houser MD, Rizzolatti G, 133-57. MIT Press.
- Dehaene S, Cohen L (2007) Cultural Recycling of Cortical Maps. *Neuron* 56:384–398.
- Dehaene S, Naccache L, Cohen L, Le Bihan D, Mangin J-F, Poline J-B, Riviere D (2001) Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience* 4:1–7.
- Dejerine J (1891) Sur un cas de cécité verbale avec agraphie suivi d'autopsie. *Mémoires de la Société de Biologie*, 43:197-201
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31:968–980.
- Démonet J-F, Chollet F, Ramsey S, Cardebat D, Nespoulous J-L, Wise RJS, Rescol A, Frackowiak R (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753–1768.
- Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology* 107:78–89.
- Dinstein I, Hasson U, Rubin N, Heeger DJ (2007) Brain Areas Selective for Both Observed and Executed Movements. *Journal of Neurophysiology* 98:1415–1427.
- Driver J (2001) A selective review of selective attention research from the past century. *British Journal of Psychology* 92:53–78.
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences* 23:475–483.
- Dykman RA, Ackerman PT (1991) Attention Deficit Disorder and Specific Reading Disability: Separate but Often Overlapping Disorders. *Journal of Learning Disabilities* 24:96–103.
- Ehri LC, Nunes SR, Willows DM, Schuster BV, Yaghoub-Zadeh Z, Shanahan T (2001) Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly* 36:250–287.
- Elliott EM, Cowan N, Valle-Inclan F (1998) The nature of cross-modal color-word interference effects. *Perception & Psychophysics* 60:761–767.
- Facoetti A, Paganoni P, Turatto M, Marzola V, Mascetti GG (2000) Visual-Spatial Attention in Developmental Dyslexia. *Cortex* 36:109–123.
- Facoetti A, Turatto M, Lorusso ML, Mascetti GG (2001) Orienting of visual attention in dyslexia: evidence for asymmetric hemispheric control of attention. *Experimental Brain Research* 138:46–53.
- Fairhall SL, Caramazza A (2013) Brain Regions That Represent Amodal Conceptual Knowledge. *Journal of Neuroscience* 33:10552–10558.
- Felton RH, Wood FB, Brown IS, Campbell SK, Harter MR (1987) Separate Verbal Memory and Naming Deficits in Attention Deficit Disorder and Reading Disability. *Brain and Language*

31:171–184.

- Fergusson DM, Horwood LJ (1992) Attention Deficit and Reading Achievement. *Journal of Child Psychology and Psychiatry* 33:375–385.
- Ferstl EC, Neumann J, Bogler C, Cramon von DY (2008) The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* 29:581–593.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RSJ, Firth CD (1995) Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition* 57:109–128.
- Frazier L, Carlson K, Clifton C Jr (2006) Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10:244–249.
- Friederici AD, Makuuchi M, Bahlmann J (2009) The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport* 20:563–568.
- Friederici AD, Meyer M, Cramon von DY (2000) Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language* 74:289–300.
- Fuentes LJ, Carmona E, Agis IF (1994) The role of the anterior attention system in semantic processing of both foveal and parafoveal words. *Journal of Cognitive Neuroscience* 6:17–25.
- Fuentes LJ, Ortells JJ (1993) Facilitation and in a Stroop-like in favor of semantic processing of parafoveally-presented stimuli interference effects task: Evidence. *Acta Psychologica* 84:213–229.
- George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM (1996) Understanding Emotional Prosody Activates Right Hemisphere Regions. 53:665–670.
- Gibson EJ, Shurcliff A, Yonas A (1970) Utilization of spelling patterns by deaf and hearing subjects. In Levin H and Williams JP (Eds.), *Basic Studies in Reading*, Basic Books, New York
- Gleitman LR, Rozin P (1977) The structure and acquisition of reading I: Relations between orthographies and the structure of language. In: Reber AS, Scarborough D (Eds.), *Toward a Psychology of Reading*. Erlbaum, Potomac, Maryland, 1–53.
- Glover GH (1999) Deconvolution of Impulse Response in Event-Related BOLD fMRI. *NeuroImage* 9:416–429.
- Golland Y, Bentin S, Gelbard H, Benjamini Y, Heller R, Nir Y, Hasson U, Malach R (2006) Extrinsic and Intrinsic Systems in the Posterior Cortex of the Human Brain Revealed during Natural Sensory Stimulation. *Cerebral Cortex* 17:766–777.
- Golumbic EMZ, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party”. *Neuron* 77:980–991.
- Gonzalez-Castillo J, Hoy CW, Handwerker DA, Robinson ME, Buchanan LC, Saad ZS,

- Bandettini PA (2015) Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *PNAS* 112:8762–8767.
- Gough PB, Hillinger ML (1980) Learning to Read: an Unnatural Act. *Bulletin of The Orton Society: An Interdisciplinary Journal of Specific Language Disability* 30:179–196.
- Graves WW, Binder JR, Desai RH, Conant LL, Seidenberg MS (2010) Neural correlates of implicit and explicit combinatorial semantic processing. *NeuroImage* 53:638–646.
- Griffis JC, Elkhetafi AS, Burge WK, Chen RH, Visscher KM (2015) Retinotopic patterns of background connectivity between V1 and fronto-parietal cortex are modulated by task demands. *Frontiers in Human Neuroscience* 9:14675.
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica* 107:293–321.
- Grimes JE (1972) *The Thread of Discourse*. Cornell University Press, Ithaca, NY.
- Hari R, Renvall H (2001) Impaired processing of rapid stimulus sequences in dyslexia. *trends in Cognitive Science* 5:525–532.
- Hasson U, Malach R, Heeger DJ (2009) Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences* 14:40–48.
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* 303:1634–1640.
- Herbster AN, Mintun MA, Nebes RD, Becker JT (1997) Regional Cerebral Blood Flow During Word and Nonword Reading. *Human Brain Mapping* 5:84–92.
- Hickok G, Poeppel D (2007) The Cortical Organization of Speech Processing. *Nat Rev Neurosci* 8:393–402.
- Hirotani M, Frazier L, Rayner K (2006) Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language* 54:425–443.
- Homae F, Hashimoto R, Nakajima K, Miyashita Y, Sakai KL (2002) From Perception to Sentence Comprehension: The Convergence of Auditory and Visual Information of Language in the Left Inferior Frontal Cortex. *NeuroImage* 16:883–900.
- Honey CJ, Thompson CR, Lerner Y, Hasson U (2012) Not Lost in Translation: Neural Responses Shared Across Languages. *Journal of Neuroscience* 32:15277–15283.
- Horton C, D'Zmura M, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology* 109:3082–3093.
- Houston SD (2004) *The First Writing*. Cambridge United Kingdom: Cambridge UP.
- Jancke L, Buchanan TW, Lutz K, Shah NJ (2001) Focused and Nonfocused Attention in Verbal and Emotional Dichotic Listening: An FMRI Study. *Brain and Language* 78:349–363.
- Jobard G, Crivello F (2003) Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage* 20:693–712.
- 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.

- Joseph JE, Cerullo MA, Farley AB, Steinmetz NA, Mier CR (2006) fMRI correlates of cortical specialization and generalization for letter processing. *NeuroImage* 32:806–820.
- Kang AM, Constable RT, Gore JC, Avrutin S (1999) An Event-Related fMRI Study of Implicit Phrase-Level Syntactic and Semantic Processing. *NeuroImage* 10:555–561.
- Kanwisher N (2010) Functional specificity in the human brain: A window into the functional architecture of the mind. *PNAS* 107:11163–11170.
- Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron* 22:751–761.
- Kayser C, Petkov CI, Augath M, Logothetis NK (2007) Functional Imaging Reveals Visual Modulation of Specific Fields in Auditory Cortex. *Journal of Neuroscience* 27:1824–1835.
- Keller TA, Carpenter PA, Just MA (2001) The Neural Bases of Sentence Comprehension: a fMRI Examination of Syntactic and Lexical Processing. *Cerebral Cortex* 11:223–237.
- Koriat A, Greenberg SN, Kreiner H (2002) The extraction of structure during reading: Evidence from reading prosody. *memory & Cognition* 30:270–280.
- Kotz SA, Meyer M, Alter K, Besson M, Cramon von DY, Friederici AD (2003) On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language* 86:366–376.
- Kouider S, de Gardelle V, Dehaene S, Dupoux E, Pallier C (2010) Cerebral bases of subliminal speech priming. *NeuroImage* 49:922–929.
- Kouider S, Dehaene S (2007) Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362:857–875.
- Kouider S, Dehaene S, Jobert A, Le Bihan D (2007) Cerebral Bases of Subliminal and Supraliminal Priming during Reading. *Cerebral Cortex* 17:2019–2029.
- Krienen FM, Yeo BTT, Buckner RL (2014) Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369:20130526–20130526.
- Kwon S, Watanabe M, Fischer E, Bartels A (2017) Attention reorganizes connectivity across networks in a frequency specific manner. *NeuroImage* 144:217–226.
- Lachter J, Forster KI, Ruthruff E (2004) Forty-Five Years After Broadbent (1958): Still No Identification Without Attention. *Psychological Review* 111:880–913.
- Lavie N, Hirst A, de Fockert JW, Viding E (2004) Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General* 133:339–354.
- Lavie N, Tsal Y (1994) Perceptual load as major determinant of the locus of selection in visual attention. *Perception & Psychophysics* 56:183–197.
- Leaver AM, Rauschecker JP (2010) Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *Journal of Neuroscience* 30:7604–7612.
- Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic Mapping of a Hierarchy of

- Temporal Receptive Windows Using a Narrated Story. *Journal of Neuroscience* 31:2906–2915.
- Levy J, Pernet C, Treserras S, Boulanouar K, Aubry F, Démonet J-F, Celsis P (2009) Testing for the Dual-Route Cascade Reading Model in the Brain: An fMRI Effective Connectivity Account of an Efficient Reading Style García AV, ed. *PLoS ONE* 4:e6675.
- Lewis JL (1970) Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology* 85:225–228.
- Lieberman AM (1992) The relation of speech to reading and writing. *Morphology Phonology. Meaning R, Frost L, Katz (Eds.), Orthography, Elsevier Science Publishers B.V, Amsterdam, 167–178.*
- Lindenberg R, Scheef L (2007) Supramodal language comprehension: Role of the left temporal lobe for listening and reading. *Neuropsychologia* 45:2407–2415.
- Lovett MW, Borden SL, DeLuca T, Lacerenza L, Bonson NJ, Brackstone D (1994) Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. *Developmental Psychology* 30:805–822.
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology* 77:24–42.
- Marinković K (2004) Spatiotemporal Dynamics of Word Processing in the Human Cortex. *The Neuroscientist* 10:142–152.
- Mattingly, IG (1972) Reading, the linguistic process, and the linguistic awareness. In: Kavanagh JF, Mattingly IG (Eds.), *Language by Ear and by Eye: The relationship between speech and reading*. MIT Press, Cambridge, MA.
- Mayberry RI, del Giudice AA, Lieberman AM (2010) Reading Achievement in Relation to Phonological Coding and Awareness in Deaf Readers: A Meta-analysis. *Journal of Deaf Studies and Deaf Education* 16:164–188.
- McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003) A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41:293–303.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748.
- Mechelli A, Crinion JT, Long S, Friston KJ, Ralph MAL, Patterson K, McClelland JL, Price CJ (2005) Dissociating Reading Processes on the Basis of Neuronal Interactions. *Journal of Cognitive Neuroscience* 17:1753–1765.
- Mesulam MM (1998) From sensation to cognition. *Brain* 121:1013–1052.
- Michael EB, Keller TA, Carpenter PA, Just MA (2001) fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping* 13:239–252.
- Miller EK, Cohen JD (2001) An Integrative Theory of Prefrontal Cortex Function. *Annual Review of neuroscience* 24:167–202.

- Mills CB, Weldon LJ (1987) Reading Text from Computer Screen. *ACM Computing Surveys* 19:329–358
- Molholm S, Foxe JJ (2005) Look “hear,” primary auditory cortex is active during lip-reading. *NeuroReport* 16:123–124.
- Moore CJ, Price CJ (1999) Three Distinct Ventral Occipitotemporal Regions for Reading and Object Naming. *NeuroImage* 10:181–192.
- Moray N (1959) Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology* 11:56–60.
- Moro A, Tettamanti M, Perani D, Donati C, Cappa SF, Fazio F (2001) Syntax and the Brain: Disentangling Grammar by Selective Anomalies. *NeuroImage* 13:110–118.
- Naccache L, Blandin E, Dehaena S (2002) Unconscious masked priming depends on temporal attention. *Psychological Science* 13:416–424.
- Narain C, Scott SK, Wise RJS, Rosen S, Leff AP, Iversen SD, Matthew PR (2003) Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI. *Cerebral Cortex* 13:1362–1368.
- National Reading Panel (2000) Report of the National Reading Panel teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instructions. Washington, DC.
- Nichols TE, Holmes AP (2001) Nonparametric Permutation Tests For Functional Neuroimaging: A Primer with Examples. *Human Brain Mapping* 15:1–25.
- Noppeney U, Price CJ (2002) A PET Study of Stimulus- and Task-Induced Semantic Processing. *NeuroImage* 15:927–935.
- Norman-Haignere SV, McCarthy G, Chun MM, Turk-Browne NB (2012) Category-Selective Background Connectivity in Ventral Visual Cortex. *Cerebral Cortex* 22:391–402.
- Nosarti C, Mechelli A, Green DW, Price CJ (2010) The Impact of Second Language Learning on Semantic and Nonsemantic First Language Reading. *Cerebral Cortex* 20:315–327.
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roetlinger M, Eulitz C, Rauschecker JP (2006) Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping* 27:562–571.
- Park H-J, Friston KJ (2013) Structural and Functional Brain Networks: From Connections to Cognition. *Science* 342:1238411–1238411.
- Paulesu E, McCloy E, Fazio F, Menoncello L, Brunswick N, Cappa SF, Cotelli M, Cossu G, Corte F, Lorusso M, Pesenti S, Gallagher A, Perani D, Price C, Frith CD, Frith U (2000) A cultural effect on brain function. *Nature Neuroscience* 3:91–96.
- Payne HE, Allen HA (2011) Active Ignoring in Early Visual Cortex. *Journal of Cognitive Neuroscience* 38:2046–2058.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics transforming numbers into movies. *Spatial Vision* 10:437–442.
- Peng D-L, Ding G-S, Perry C, Xu D, Jin Z, Luo Q, Zhang L, Deng Y (2004) fMRI evidence for

- the automatic phonological activation of briefly presented words. *Cognitive Brain Research* 20:156–164.
- Perfetti CA, Sandak R (2000) Reading Optimally Builds on Spoken Language: Implications for Deaf Readers. *Journal of Deaf Studies and Deaf Education* 5:32–50.
- Posner MI, Rothbart MK (1998) Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences* 353:1915–1927.
- Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences* 1191:62–88.
- Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62:816–847.
- Price CJ, Devlin JT (2003) The myth of the visual word form area. *NeuroImage* 19:473–481.
- Price CJ, Devlin JT (2011) The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences* 15:246–253.
- Price CJ, Mechelli A (2005) Reading and reading disturbance. *Current Opinion in Neurobiology* 15:231–238.
- Price CJ, Price CJ, Wise RJS, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RSJ, Friston KJ (1996a) Hearing and saying. *Brain* 119:919–931.
- Price CJ, Wise RJS, Frackowiak RSJ (1996b) Demonstrating the Implicit Processing of Visually Presented Words and Pseudowords. *Cerebral Cortex* 6:62–70.
- Protzner AB, McIntosh AR (2007) The Interplay of Stimulus Modality and Response Latency in Neural Network Organization for Simple Working Memory Tasks. *Journal of Neuroscience* 27:3187–3197.
- Pulvermüller F, Shtyrov Y (2006) Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology* 79:49–71.
- Purvis KL, Tannock R (1997) Language Abilities in Children with Attention Deficit Hyperactivity Disorder, Reading Disabilities, and Normal Controls. *Journal of Abnormal Child Psychology* 25:133–144.
- Raij T, Uutela K, Hari R (2000) Audiovisual Integration of Letters in the Human Brain. *Neuron* 28:617–625.
- Rama P, Relander-Syrjänen K, Carlson S, Salonen O, Kujala T (2012) Attention and semantic processing during speech: An fMRI study. *Brain and Language* 122:114–119.
- Rayner K (1978) Eye Movements in Reading and Information Processing. *Psychological Bulletin* 85:618–660.
- Rees G, Russell C, Frith CD, Driver J (1999) Inattentional Blindness Versus Inattentional Amnesia for Fixated But Ignored Words. *Science* 286:2504–2507.
- Richardson FM, Thomas MSC, Price CJ (2010) Neuronal Activation for Semantically Reversible Sentences. *Journal of Cognitive Neuroscience* 22:1283–1298.
- Romanski LM (2007) Representation and Integration of Auditory and Visual Stimuli in the

- Primate Ventral Lateral Prefrontal Cortex. *Cerebral Cortex* 17:i61–i69.
- Romanski LM, Averbeck BB (2009) The Primate Cortical Auditory System and Neural Representation of Conspecific Vocalizations. *Annual Review of neuroscience* 32:315–346.
- Ross ED, Mesulam M-M (1979) Dominant Language Functions of the Right Hemisphere? *Arch Neurol* 36:144–148.
- Ruz M, Wolmetz ME, Tudela P, McCandliss BD (2005) Two brain pathways for attended and ignored words. *NeuroImage* 27:852–861.
- Sabri M, Binder JR, Desai R, Medler DA, Leitt MD, Liebenthal E (2008) Attentional and linguistic interactions in speech perception. *NeuroImage* 39:1444–1456.
- 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.
- Schmithorst VJ, Holland SK, Plante E (2006) Cognitive modules utilized for narrative comprehension in children: a functional magnetic resonance imaging study. *NeuroImage* 29:254–266.
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, “unisensory” processing. *Current Opinion in Neurobiology* 15:454–458.
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences* 32:9–18.
- Seghier M, Price CJ (2011) Explaining Left Lateralization for Words in the Ventral Occipitotemporal Cortex. *Journal of Neuroscience* 31:14745–14753.
- Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage* 44:223–231.
- Serences JT, Yantis S, Culbertson A, Awh E (2004) Preparatory Activity in Visual Cortex Indexes Distractor Suppression During Covert Spatial Orienting. *Journal of Neurophysiology* 92:3538–3545.
- 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.
- Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2011) Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex* 22:158–165.
- Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *PNAS* 111:E4687–E4696.
- Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U (2016) Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications* 7:1–13.
- Smith EE, Jonides J (1999) Storage and Executive Processes in the Frontal Lobes. *Science* 283:1657–1661.

- Snowling MJ, Hulme C (2005) *The Science of Reading: A handbook*. Malden, MA: Blackwell.
- Spadone S, Penna Della S, Sestieri C, Betti V, Tosoni A, Perrucci MG, Romani GL, Corbetta M (2015) Dynamic reorganization of human resting-state networks during visuospatial attention. *PNAS* 112:8112–8117.
- Specht K, Reul J (2003) Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *NeuroImage* 20:1944–1954.
- Spitsyna G, Warren J, Scott SK (2006) Converging Language Streams in the Human Temporal Lobe. *The Journal of Neuroscience* 26:7328–7336.
- Stein von A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha. *International Journal of Psychophysiology* 38:301–313.
- Stroop JR (1935) Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology* 18:643–662.
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America* 26:212–215.
- Swets B, Desmet T, Hambrick DZ, Ferreira F (2007) The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General* 136:64–81.
- Talairach J, Tournoux P (1988) *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers.
- Thierry G, Giraud A-L, Price CJ (2003) Hemispheric Dissociation in Access to the Human Semantic System. *Neuron* 38:499–506.
- Treisman AM (1960) Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology* 12:242–248.
- Treisman AM, Squire R, Green J (1974) Semantic processing in dichotic listening? A replication. *memory & Cognition* 2:641–646.
- Vagharchakian L, Dehaene-Lambertz G, Pallier C, Dehaene S (2012) A Temporal Bottleneck in the Language Comprehension Network. *The Journal of Neuroscience* 32:9089–9102.
- 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.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2004) Integration of Letters and Speech Sounds in the Human Brain. *Neuron* 43:271–282.
- van Dijk T, Kintsch W (1983) *Strategies of Discourse Comprehension*. New York: Academic.
- Vandenberghe R, Price CJ, Wise RJS, Josephs O, Frackowiak RSJ (1996) Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254–256.
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-

- Form System. *Neuron* 55:143–156.
- Wagner RK, Torgesen JK (1987) The Nature of Phonological Processing and Its Causal Role in the Acquisition of Reading Skills. *Psychological Bulletin* 101:192–212.
- Wang M, He BJ (2014) A cross-modal investigation of the neural substrates for ongoing cognition. *Frontiers in Psychology* 5:1–10.
- Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H (2005) Identification of emotional intonation evaluated by fMRI. *NeuroImage* 24:1233–1241.
- Wilson SM, Molnar-Szakacs I, Iacoboni M (2007) Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension. *Cerebral Cortex* 18:230–242.
- Wise RJS, Scott SK, Blank C, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within “Wernicke” sarea’. *Brain* 124:83–95.
- Woollams AM, Silani G, Okada K, Patterson K, Price CJ (2010) Word or Word-like? Dissociating Orthographic Typicality from Lexicality in the Left Occipito-temporal Cortex. *Journal of Cognitive Neuroscience* 23:992–1002.
- 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.
- Yeung AS, Jin P, Sweller J (1997) Cognitive Load and Learner Expertise: Split-Attention and Redundancy Effects in Reading with Explanatory Notes. *Contemporary Educational Psychology* 23:1–21.
- Zarahn E, Aguirre GK, D’Esposito M (1997) Empirical Analyses of BOLD fMRI Statistics. *NeuroImage* 5:179–197.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of Phonetic and Pitch Discrimination in Speech Processing. *Science* 256:846–849.
- Zwaan RA, Radvansky GA (1998) Situation Models in Language Comprehension and Memory. *Psychological Bulletin* 123:162–185.