

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

How does inattention affect written and spoken language processing?



Laurent Cohen^{a,b,*}, Philippine Salondy^a, Christophe Pallier^c and Stanislas Dehaene^{c,d}

^a Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau, ICM, Paris, France

^b AP-HP, Hôpital de La Pitié Salpêtrière, Fédération de Neurologie, Paris, France

^c Cognitive Neuroimaging Unit, CEA DRF/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, Gif/Yvette, France

^d Collège de France, Paris, France

ARTICLE INFO

Article history:

Received 27 October 2020

Reviewed 15 December 2020

Revised 20 December 2020

Accepted 2 February 2021

Action editor Sonja Kotz

Published online 19 February 2021

Keywords:

Reading

Language

Attention

fmri

ABSTRACT

The classic cocktail party effect suggests that some, but probably not all levels of language processing can proceed without attention. We used whole-brain functional MRI to investigate how modality-specific and modality-independent language areas are modulated by the withdrawal of attention to another sensory modality (e.g., attending to vision during the presentation of auditory sentences, or vice-versa). We tested the hypotheses that inattention may abolish sentence-level integration and eliminate top-down effects. In both written and spoken modalities, language processing was strongly modulated by the distraction of attention, but this inattention effect varied considerably depending on the area and hierarchical level of language processing. Under inattention, a bottom-up activation remained in early modality-specific areas, particularly in superior temporal spoken-language areas, but the difference between sentences and words lists vanished. Under both attended and unattended conditions, ventral temporal cortices were activated in a top-down manner by spoken language more than by control stimuli, reaching posteriorly the Visual Word Form Area. We conclude that inattention prevents sentence-level syntactic and semantic integration, but preserves some top-down crossmodal processing, plus a large degree of bottom-up modality-specific processing, including a ventral occipito-temporal specialization for letter strings in a known alphabet.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author. Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau, ICM, F-75013, Paris, France.

E-mail address: laurentcohen2@gmail.com (L. Cohen).

<https://doi.org/10.1016/j.cortex.2021.02.007>

0010-9452/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

What happens when we stop attending to a language stream? Which aspects of language processing occur without attention? Which are modulated by top-down attentional processes? The well-known Cocktail Party Problem refers to the fact that we continue to recognize what one speaker is saying when another is talking at the same time, thus questioning the fate of unattended speech (Bronkhorst, 2015; Cherry, 1953; Hill & Miller, 2009; McDermott, 2009). There is evidence that attention modulates speech processing at all levels ranging from the brainstem (Forte et al., 2017) to the auditory cortex (Mesgarani & Chang, 2012) and high-level language areas (Sabri et al., 2008). However, although the role of attention in normal and impaired reading acquisition has been well established (Facoetti et al., 2006; Grainger et al., 2016; Lobier et al., 2012), its interaction with brain reading mechanisms has received relatively little scrutiny, as compared to speech perception (Cohen et al., 2008; Pattamadilok et al., 2017).

Here, our aim was to test the effects of inattention on sentence-level processing. Can we still integrate multiple words into a syntactically and semantically coherent structure, without attending to those words? Since the inception of brain imaging, the contrast between processing sentences and word lists has been used to demonstrate the existence of a network of brain areas that activate more when multiple words can be unified (Fedorenko & Thompson-Schill, 2014; Goucha & Friederici, 2015; Mazoyer et al., 1993; Moreno et al., 2018; Nelson et al., 2017; Pallier et al., 2011). The relevant areas include the left inferior frontal region (“Broca’s area”) and a large extent of the superior temporal sulcus, extending anteriorly in the temporal pole and posteriorly in the angular gyrus.

Here, our first question is to what extent this network continues to respond more to sentences than to word lists when attention is entirely withdrawn from the stimuli, by asking subjects to perform a difficult task on concurrent stimuli presented in another modality. Since the degree of automaticity could conceivably be different for spoken language, which has been subject to a long evolution, and for written language, which is a recent cultural invention, we present sentences and word lists using both spoken and written words, and ask whether the sentence > word lists effect identically disappears under inattention in both cases.

Our second goal was to investigate, more specifically, the role of attention on top-down processing in ventral visual cortex during language processing. This aspect of our study is part of a longer-term research program on the role of ventral occipitotemporal cortex (VOT) in word reading, and on the bottom-up and top-down influences which drive its function. Hence we particularly examined how attention modulates the VOT activation induced in a bottom-up manner by printed stimuli, and in a top-down manner by spoken stimuli.

Those questions are motivated by an unsolved theoretical debate concerning the role of top-down inputs in the processing of written language in the left ventral occipitotemporal visual pathway, including the classical site of the visual word form area (VWFA, Cohen et al., 2000; Dehaene et al., 2015; Taylor et al., 2019), a region that shows a

preference for alphabetic stimuli over other types of images, and consists of cortical patches with various sensitivity to lexical frequency, graphemic structure, case invariance, etc. (for reviews, see e.g., Bouhali et al., 2019; Dehaene et al., 2015). According to the bottom-up hypothesis, the VWFA owes its selectivity for written stimuli to a series of bottom-up stages that progressively compute an invariant representation of written letter strings (Dehaene et al., 2005). According to the alternative top-down hypothesis, the VWFA’s apparent selectivity is entirely due to top-down signals from higher-level language areas that transiently recruit this site into a broader network for the purpose of reading (Carreiras et al., 2014; Price and Devlin, 2003, 2011; Whaley et al., 2016; Woodhead et al., 2014).

Note that both sides of the debate agree about the existence of top-down effects at this site. fMRI has demonstrated that the VWFA, although belonging to the ventral visual pathway, can be activated in a top-down during speech perception, for instance when subjects attend to syllables versus tones (Yoncheva et al., 2010). The top-down effect is more intense when participants are literate (Dehaene et al., 2010), but may already be present in babies (Dehaene-Lambertz et al., 2018; Li et al., 2019), suggesting that the VWFA may owe its specific location to the existence of an early innate connectivity pattern which makes this region part of the left-hemispheric language network (Barttfeld et al., 2018; Bouhali et al., 2014; Hannagan et al., 2015; Saygin et al., 2012). In adults, however, such top-down activation seems to occur only when orthographic access is beneficial to task performance, e.g., in auditory lexical decision, but not when it is task-irrelevant, e.g., during simple sentence comprehension (Dehaene et al., 2010). The theoretical disagreement does not concern the existence of top-down effects, but whether they suffice to explain the selectivity for written words in the VWFA. We reasoned that, if inattention entirely withdraws all top-down resources, then measuring the remaining activation should help separating the two hypotheses: according to the top-down hypothesis, there should no longer be any selectivity for written words relative to control stimuli (here, stimuli in another alphabet, unreadable by our fMRI participants), whereas according to the bottom-up hypothesis, such selectivity should remain partially present.

To address those questions, our experiment examined the effects of inattention on spoken and written language processing. In separate fMRI blocks, we presented sentences, word lists, and non-linguistic control stimuli (visual: another alphabet; auditory: temporally scrambled words). The other experimental factors were stimulus modality (visual or auditory) and attention (attended versus unattended linguistic stream). In the unattended condition, to maximize inattention, participants were fully engaged in a difficult task in the other modality (see Fig. 1).

Anticipating on the results, we replicated the classical network of left-lateralized temporal and inferior frontal regions, more activated by sentences than by word lists. We found that this language network is massively modulated by inattention, to such an extent that all differences between sentences and word lists vanish when the stimulus modality ceases to be attended. However, we also found a preservation of modality-specific processing for all language conditions

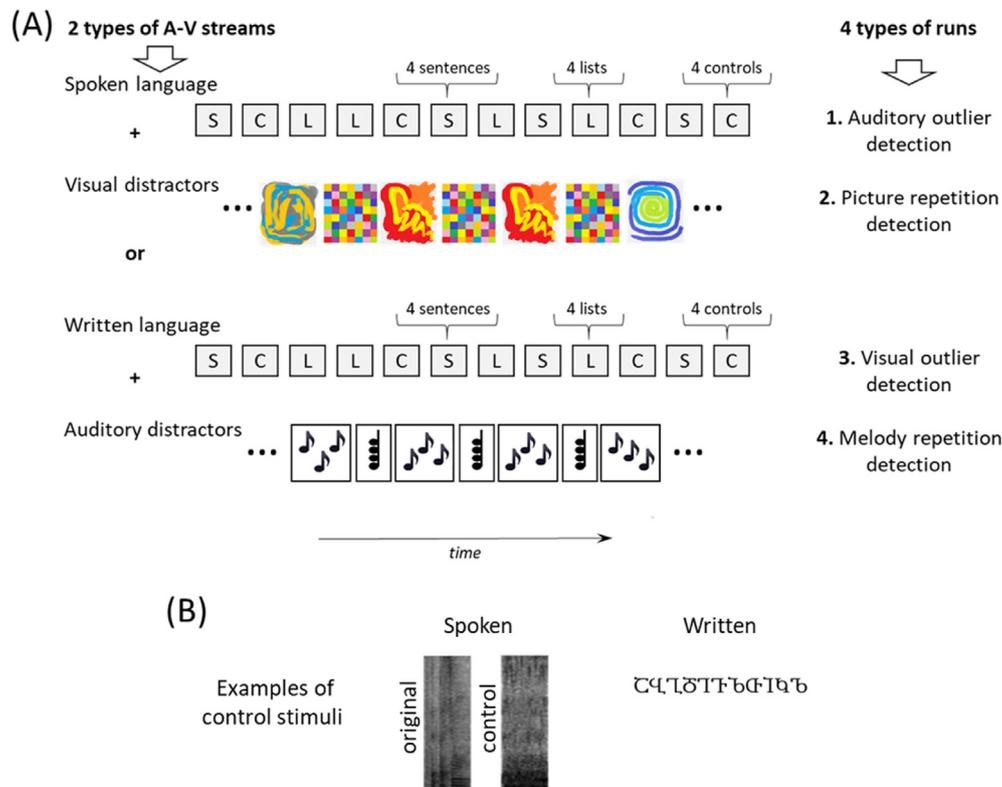


Fig. 1 – Summary of the experimental design. (A) Participants were presented simultaneously, in two different modalities, with a stream of language stimuli and with another unsynchronized stream of distractors. Language stimuli could be spoken (top panel) or written (bottom panel), and distractors were always in the opposite modality. Language stimuli consisted in blocks of 4 sentences [S], 4 lists of words [L], or 4 sequences of low-level controls [C]. Participants were asked either to perform a target detection task on language stimuli (attended condition), or to detect repetitions of distractors (unattended condition), resulting in 4 types of language modality x task runs. **(B)** Low-level controls consisted in scrambled speech, as illustrated here by sample spectrograms (Joly et al., 2012), and in strings of letters in the unknown Georgian alphabet.

relative to their non-linguistic controls. Thus, a selective activation of left occipito-temporal cortex to written words continued to be present even when those stimuli were unattended. Furthermore, the top-down activation of the ventral temporal cortex by auditory stimuli was reduced but persisted under inattention, with the VWFA standing at the posterior tip of this top-down activation.

2. Methods

Study data and digital study materials are freely available at <https://doi.org/10.5281/zenodo.4529179>. No part of the study procedures and analyses was pre-registered prior to the research being conducted. We report all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. General design

Throughout the experiment, participants were presented simultaneously with a stream of language stimuli, and with a

stream of distractors. We used a crossed experimental design with three within-subject factors (Fig. 1). The first factor was modality: Language stimuli could be either written or spoken, while distractors were always in the opposite modality that is auditory or visual, respectively. The second factor was structure: the stimuli consisted in sentences, lists of words, or low-level controls. The third factor was attention: In the attended condition, participants were asked to perform a target detection task on language stimuli, while in the unattended condition they had to disregard language stimuli while performing a repetition detection task with on distractors.

2.2. Participants

Sixteen right-handed native French speakers (8 women, 22.2 ± 1.8 years old) with higher education participated in the study and provided written informed consent. The experiment was approved by the institutional review board of the INSERM (protocol C13-41). Participants had no history of dyslexia, nor of neurological or psychiatric disorders. They all had normal or corrected to normal vision.

2.3. Stimuli

2.3.1. Sentences and lists

We created 40 12-word sentences, and derived from each sentence 7 other sentences matched in syntactic structure, yielding a total of 8 parallel sets of 40 sentences. We checked that open-class words (common nouns, adjectives, non-auxiliary verbs, adverbs) did not differ across the sets in frequency (as measured by log lemma frequency, www.lexique.org), in number of syllables, and in number of letters. We swapped some sentences between sets to further minimize differences. Within each of the 8 sets (but not within sentences), words were then randomly shuffled so as to create 40 12-word lists with no syntactic structure. Lists were checked manually, and words were permuted within or between lists whenever necessary to avoid accidental syntactic structures or lists with too long or too short duration. Each of the 8 set of 40 stimuli thus existed in 4 versions: sentence or list, and written or spoken.

2.3.2. Low-level controls

One set of word lists was used to generate 40 visual and 40 auditory low-level control stimuli. Visual controls were created by replacing the original letters by letters from the Georgian alphabet, unknown to the participants, using a visually comparable font. Auditory controls were created by scrambling temporally the original speech stimuli, separately within a series of frequency bands (Joly et al., 2012).

2.3.3. Targets

From each stimulus we derived a second version in which a target was introduced. In sentences and lists, targets consisted in a pseudoword replacing one of the open-class words. The pseudoword comprised the same number of letters and syllables as the original word. In low-level visual controls, targets consisted in a string of a repeated letter from the Georgian alphabet, with the same number of letters as the original string which it replaced. In low-level auditory controls, targets consisted in a set of 40 sounds with the same average duration as real words (610 msec), made up of 4 repetitions of a brief fragment (mean 136 msec) of phase-shifted speech. They could be spotted out in the stream of phase-shifted speech on the basis of their repeated acoustic pattern. Targets occurred equally often in the first, second, and last third of stimuli.

2.3.4. Duration

Auditory sentences were temporally scaled by a factor of 1.15, lists by a factor of .88, and auditory controls by a factor of .73, in order to equate the average duration of all types of spoken stimuli. We checked that the 8 sets of stimuli did not differ in duration ($F(1,7) = 1.18; p = .3$). Auditory stimuli were synthesized using the Balabolka speech software (balabolka.fr.softonic.com/) and in-house scripts. In written stimuli, the 12 words were displayed successively for .45 sec each, for a fixed total duration of 5.4 sec, which was the mean duration of spoken stimuli.

2.3.5. Distractors

Visual distractors consisted in 30 abstract color drawings. Auditory distractors consisted in 30 melodies of 3 notes. Distractors were displayed for 600 msec, followed by a 400 msec mask. The visual mask consisted in a colored checkerboard covering the area of stimuli. The auditory mask consisted in a chord compounding many simultaneous notes.

2.4. Procedure

As described before, we generated 8 sets of 40 stimuli, each set existing in 4 versions (sentences/lists x written/spoken), plus one set of 40 low-level controls existing in 2 versions (written/spoken). Moreover, each of those stimuli existed in 2 versions, one with a target and one without target.

2.4.1. Assignment of stimuli to participants

For each participant, the 8 sets of language stimuli were randomly assigned to the 8 conditions consisting in the combinations of written/spoken x sentences/lists x attended/unattended. Within each condition, stimuli were randomly split into one half with a target, and one half without a target. Thanks to this procedure, no sentence or list was ever presented more than once to a given subject, avoiding any potential repetition suppression or habituation. For each participant and each condition, 32 stimuli were drawn randomly from the appropriate set of 40 items, shuffled randomly, and inserted in the sequence of experimental events, as described below.

2.4.2. Temporal structure of the experiment

The experiment consisted in 8 runs, two for each of the 4 combinations of written or spoken language and attended or unattended stimuli. The order of the runs was fixed. The first 4 runs were: spoken attended, written attended, written unattended, spoken unattended. The last 4 runs followed the reverse order, forming an embedded ABBA design. Runs had a mean duration of 7 min and 4 sec.

Each run started with instructions displayed visually for 5 sec. Instructions specified whether participants had to pay attention to language stimuli and detect targets or to distractors and perform a one-back repetition detection task. Instructions were followed by a 15 sec rest period. Then, the stream of distracting stimuli started. Distracting stimuli were presented alone during 15 sec, and then the presentation of language stimuli started. After the last language stimulus in the run, there was again a 15 sec period during which only the distractors were presented, followed by a 15 sec rest period.

Language stimulation consisted in a random alternation of short blocks of sentences, lists, or low-level control stimuli. Each short block comprised 4 stimuli of the same type. Each run included 4 blocks (i.e., 16 stimuli) of each of the 3 structure types. Half the stimuli included a target. Within blocks, each stimulus was followed by a 1.5 sec rest period, except for the last one, which was followed by a 5 sec rest period.

Written stimuli were printed in black on a light gray background. A black central fixation cross was present during all rest periods.

Distractors were presented in random order during 600 msec, alternating with the 400 msec mask. About ¼ of distractors (24% for vision and 27% for audition) were repeated twice in a row.

2.4.3. Tasks

During attended blocks, participants had to press a button when detecting a pseudoword in sentences or lists, a string containing repeated letters in visual controls, and a sound pattern consisting of repeated sounds in auditory controls. During the unattended blocks, participants had to press a button when detecting a repeated distractor (one-back task). Half the subjects responded with their right hand during the first 4 blocks and their left hand during the last 4 blocks, and conversely for the other half of subjects.

2.5. MRI acquisition and analysis

2.5.1. Acquisition

Participants were scanned using a 3T MRI scanner (Siemens, PRISMA) with a 64-channel head coil. Functional data were acquired with a triple gradient echo multi-band sequence sensitive to brain oxygen-level-dependent (BOLD) contrast (54 axial slices, 2.5 mm thickness, TR = 1600 msec, flip angle = 73°, TE1 = 15.2 msec, TE2 = 37.17 msec, TE3 = 59.14 msec, in-plane-resolution = 2.5 × 2.5 mm, matrix = 84 × 84, acceleration factor in plane iPat = 2 and slice multiband = 3). The triple echo sequence was chosen in order to decrease signal drop-out, especially in ventral regions located above the ear canals, hence sensitive to susceptibility artifacts, and to increase signal-to-noise ratio (Poser et al., 2006). For each run, additional BOLD volumes with reverse phase encoding direction were also acquired. T1-weighted images were acquired for anatomical localization (192 axial slices, 1 mm thickness, TR = 2300 msec, flip angle = 9°, TE = 2.76 msec, TI = 900 msec, in-plane-resolution = 1 × 1 mm, matrix = 256 × 256, acceleration factor in plane iPat = 2).

2.5.2. Preprocessing

Images were preprocessed with MultiEcho-ICA (ME-ICA) v2.5 (Kundu et al., 2012, 2013) (<https://github.com/ME-ICA/me-ica>) and AFNI (Cox, 1996). They were despiked, slice-timing corrected, motion corrected, and coregistered to the bias-corrected and skull-stripped anatomical volume, and cleaned from thermal and physiological noise. Motion correction and coregistration were computed using the first echo, and applied to the other two echoes. The 3 echoes were then combined using a weighted average, and an ICA was performed to reduce data dimensionality by removing components related to physiological noise, while keeping BOLD components. Anatomical and functional images were normalized and functional images were smoothed using a Gaussian kernel of 4 mm full-width at half maximum (FWHM).

2.5.3. First-level GLM and second-level analyses

Analyses were performed using the SPM12 toolbox for Matlab. Each participant's functional data were separated into 8 runs

and combined into one single GLM model. For each run, we defined 6 regressors for experimental conditions (sentences/lists/controls × first/second half of stimuli), plus a regressor for button presses, and one modeling the onset and offset of the stream of distractors. Regressors were convolved with the canonical SPM hemodynamic response function, in addition to six motion parameters. For each subject, we computed the contrast image for each combination of modality × sentences/lists/controls × attention × first/second half of stimuli. Those individual images were smoothed (6 mm FWHM Gaussian filter), masked by a precomputed voxel-mask containing voxels that were scanned in half of the participants or more, and entered in a second-level ANOVA model with subjects as random factor.

All results reported were obtained using a voxelwise statistical threshold of $p < .001$, and a clusterwise threshold of $p < .05$ corrected for multiple comparisons over the whole brain, or within a left VOT region of interest whenever specified. Correction for multiple comparisons was based on Random Field Theory as implemented in the SPM12 software (Friston et al., 1994; Nichols, 2012). This region of interest was defined by combining the left inferior occipital, inferior temporal, and fusiform gyri from the AAL atlas, truncated between $y = -65$ and $y = -25$ (Supplementary Figure 1).

2.5.4. Modeling sentence processing

Contrary to the perception of unstructured lists of words, the processing of sentences is not a steady process. Some core computations leading to the syntactic and semantic closure of complex structures are carried out only by the end of sentences (Pallier et al., 2011). In order to determine the contrast most sensitive to sentence-level structure, we used distinct regressors for the first and second halves of sentences (see above), and examined activations induced by the first and the second half of sentences, relative to the corresponding halves of word lists. With both visual and auditory language, the first half of sentences yielded no significant activation at the usual threshold, while the second half of sentences activated strongly the left frontotemporal sentence-level network, as described in the Results. Hence, when comparing language stimuli (sentences or word lists) with low-level controls, we used a contrast that bore on both the 1st and 2nd halves of stimuli. However, when comparing sentences and word lists, we compared only the activations evoked by the second half of sentences and lists.

3. Results

3.1. Behavioral results

Hits were defined as button presses occurring in the interval between 300 msec after target onset and 2 sec after target ending. Individual percent hits were computed for each condition and entered in ANOVAs with subjects as random factor (Supplementary Figure 1). Participants made at least 88% hits in all conditions, except when detecting pseudowords in spoken sentences (61% hits) and in spoken lists (46% hits). Each of those two conditions was significantly below the other four language conditions (both $p < .001$), which did not differ

($p > .05$). False alarms followed a symmetrical pattern: they were below 8% in all conditions, except in the two conditions with a relatively low hit rate (Supplementary Figure 1).

For the detection of repeated auditory and visual distractors, performance was around 90%, and did not differ between the two modalities ($F < 1$). Such high performance implies that attention manipulation allowed non-language distractors to receive high processing resources whenever we wished to direct attention away from language.

In summary, performance was excellent except with spoken language. The latter result is probably due to the difficulty of hearing the difference between real words and the rare target pseudowords in the noisy environment of the magnet, which was not the case for written stimuli. Nevertheless this difference is not essential to the goals of the present study, since (1) behavioral evidence indicated that participants were fully and equally engaged in the distracting tasks with auditory and visual stimuli; (2) as we shall now see, there was ample fMRI evidence for a difference between spoken sentences versus word lists, indicating that the auditory stimuli were clearly perceived and processed. If anything, our experimental condition maximized the difference between attended and unattended conditions.

3.2. Imaging results

3.2.1. Whole-brain activations

We first study the whole-brain activations evoked by attended stimuli. In each analysis, we first assess the conjunction of activations by written and spoken stimuli, and then the differences between modalities.

3.2.2. Attended language

The conjunction of written and spoken language (averaging sentences and lists) minus their respective controls (voxelwise $p < .001$ each; Fig. 2A) activated a set of left-hemispheric supramodal language regions: the anterior fusiform gyrus (MNI -45 -34 -21, $Z = 5.81$) anterior to the VWFA proper, the whole extent of the superior temporal sulcus (STS; MNI -54 -47 4, $Z > 8$; MNI -52 -12 -8, $Z = 6.29$), the superior temporal pole (MNI -52 13 -13, $Z = 5.33$), Broca's area (MNI -40 28 -3, $Z = 6.56$), the precentral gyrus (MNI -45 15 24, $Z = 7.49$), and the SMA (MNI -5 8 63, $Z = 7.60$). Activation for all experimental conditions at the main peaks are displayed in Supplementary Figure 2.

The vast majority of these activations did not differ between modalities. The contrast of language minus control, for spoken minus for written language (masked by the same contrast for spoken stimuli, voxelwise $p < .01$) showed only bilateral STG activations (Fig. 2A right panel; left: MNI -62 -15 2, $Z = 7.61$; right: MNI 62 -10 -6, $Z > 8$). The opposite contrast of written minus spoken language showed no significant activation.

3.2.3. Attended sentences

The conjunction of written and spoken sentences minus the corresponding word lists (voxelwise $p < .001$ each; Fig. 2B) activated the left superior temporal sulcus (MNI -50 -12 -13, $Z = 4.71$), temporal pole (MNI -47 13 -26, $Z = 6.03$), and Broca's area (MNI -42 28 -11, $Z = 4.86$; MNI -54 25 19, $Z = 4.59$), and the

anterior left VOT (MNI -42 -19 -23, $Z = 3.71$, corrected in the left VOT ROI). The interaction contrast (sentences $>$ lists) \times (written $>$ spoken stimuli), masked by written sentences minus rest (voxelwise $p < .01$), showed the same regions, indicating that most of the supramodal sentence-specific network was more activated for written than spoken stimuli (Fig. 2B right panel). The opposite contrast for spoken minus written stimuli showed no significant activation.

3.2.4. Effect of inattention on language processing

We next examined whether the difference between language and control stimuli was larger under attended than under unattended conditions (Fig. 3A, left panel). This contrast, in the conjunction of written and spoken stimuli, showed that in both modalities, most of the left-hemispheric supramodal language network was significantly less activated when language was unattended, including the STS (MNI -52 -47 4, $Z = 4.44$), Broca's area (MNI -40 28 -1, $Z = 4.69$; MNI -47 15 24, $Z = 5.39$), the precentral gyrus (MNI -45 0 51, $Z = 5.06$) and the SMA (MNI 5 -5 63, $Z = 6.68$).

Under inattention, the remaining activation by language minus controls included (Fig. 3B), for spoken stimuli, bilateral STS/STG (MNI -59 -10 -3, $Z > 8$; MNI 59 -5 -8, $Z > 8$) and anterior VOT; and, for written stimuli, bilateral mesial occipital cortex (MNI -5 -94 21, $Z = 4.82$; MNI 10 -86 19, $Z = 4.57$), and left anterior fusiform (MNI -42 -37 -18, $Z = 4.67$).

The interaction contrast of (spoken $>$ written) \times (attended $>$ unattended) language, masked by spoken attended language (voxelwise .01), and the opposite contrast of written minus spoken language, showed no significant activation. In other terms, there was no significant difference between modalities in the effect of attention on overall language processing.

3.2.5. Effect of inattention on sentences versus word lists

The conjunction of the effect of inattention on (sentences $>$ lists) for written and for spoken stimuli, showed no activation. This was because the effect of inattention on (sentences $>$ lists) did not reach significance for spoken language, while there was a massive effect of inattention for written language in the entire sentences>lists network (Fig. 3A right panel). When the threshold was lowered to voxelwise .01, however, a similar effect of inattention emerged with spoken stimuli in the temporal poles and Broca's area. Thus, inattention, in either modality, reduced sentence-level processing, though this effect was more stable with written stimuli. Importantly, in both modalities, not a single voxel showed a significant difference between sentences and word lists when the stimuli were unattended, suggesting that sentence-level processing is drastically reduced when attention is removed. Additional Bayesian analyses of the contrast between sentences and lists under inattention supported this conclusion (see Supplementary Figure 3).

Having found strong effects of all experimental factors at the whole-brain level, we now focus on a detailed study of activation patterns in the left VOT.

3.3. Activation in the left VOT cortex

Results are depicted in Figs. 4 and 5.

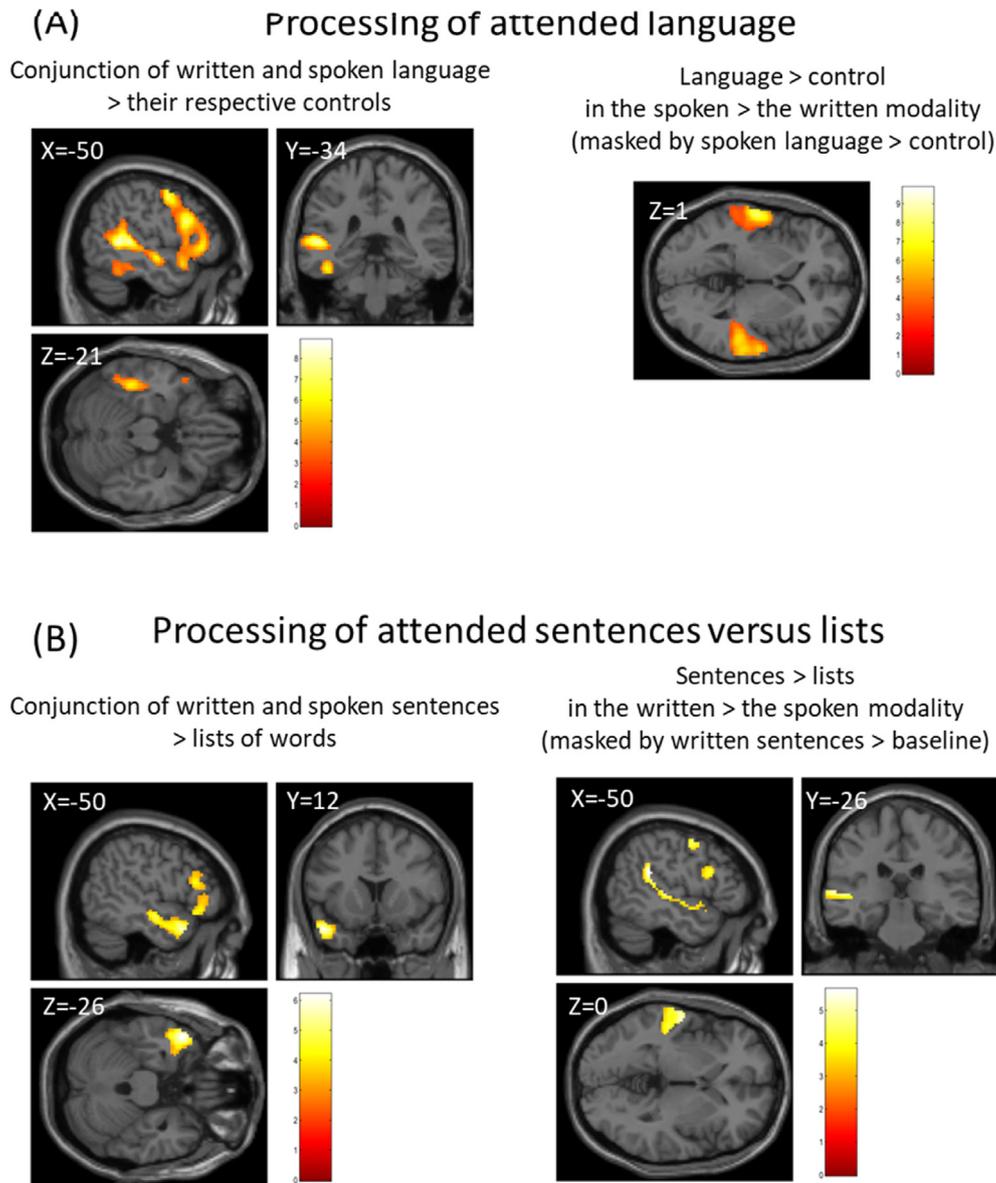


Fig. 2 – Whole-brain analyses of the activations evoked by language when it is attended. Activations common to both spoken and written language appear in the left column, and differences between modalities in the right column. (A) Contrast of language (averaging over sentences and word lists) minus control. (B) Contrast of sentences minus lists of words.

3.4. Activation by written language

3.4.1. Attended written language

Relative to baseline, written language as well as control stimuli activated the VOT cortex extensively (Fig. 4, top row). There was a cross-over pattern, with stronger activation for control stimuli than for written language posterior to about $Y = -60$, and the converse anterior to this boundary. Specifically, significant activations by language > control peaked just anterior to the usual coordinates of the VWFA (MNI -45 -34 -21, $Z = 5.81$), extending from $Y = -52$ to $Y = -20$. This activation overlapped with the supramodal activation reported above in whole-brain

analyses ($Y = -34$). The usual peak of the VWFA (MNI -42 -57 -15; Cohen et al., 2002), fell close to the cross-over point: it was strongly activated by both control and language stimuli ($Z > 8$ for both), which did not differ ($Z = -.92$).

3.4.2. Attended written sentences

The contrast of sentences minus lists showed anterior VOT activation extending from the inferior temporal pole (MNI -42 -10 -33, $Z = 4.75$) back into the anterior fusiform gyrus (MNI -42 -37 -23, $Z = 3.71$), including the peak for attended written language reported just before (MNI -45 -34 -21, $Z = 3.51$) (Fig. 4, top row).

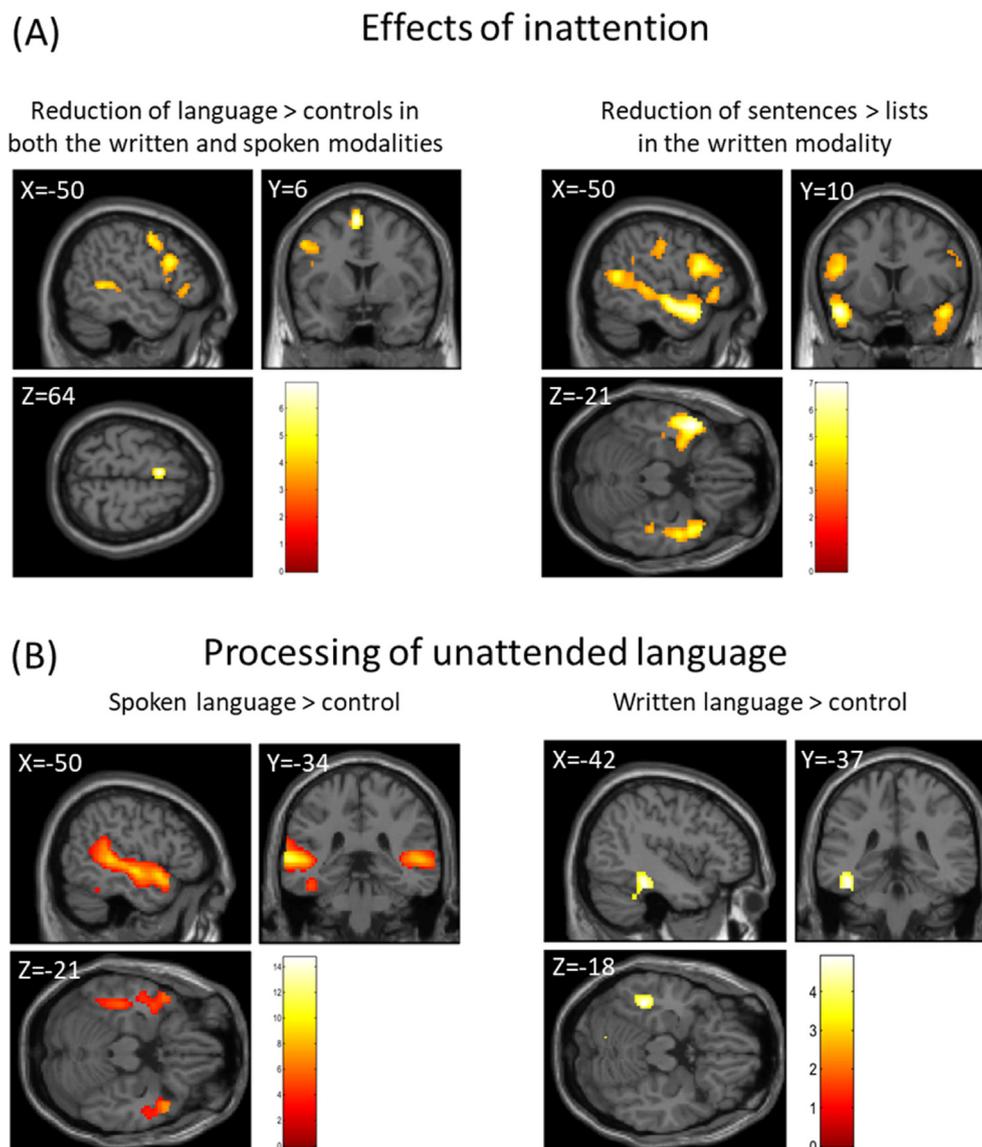


Fig. 3 – Whole-brain analyses of the effect of inattention on language processing. (A) Reduction in activation by unattended as compared to attended stimuli. Left: By both spoken and written language minus controls. Right: By written sentences minus lists of words. (B) Residual activation by unattended stimuli. Left: By spoken language minus control. Right: By written language minus control.

3.4.3. Effect of attention on written language processing

The contrast of attended > unattended stimuli showed that removal of attention decreased activation by both language and control in most areas, while preserving their relative relationships, including the cross-over around $Y = -60$ (Fig. 4, middle row). Specifically, the effect of inattention on language activations relative to baseline ranged from the occipital tip (MNI -22 -91 -6, $Z = 5.28$) to the anterior fusiform (MNI -45 -34 -18, $Z = 4.07$), including the usual VWFA (MNI -42 -57 -8, $Z = 4.64$). Although the cross-over pattern was unchanged, inattention weakened the contrast of control > language in posterior regions (MNI -27 -62 -11, $Z = 4.84$; cluster ranging from $Y = -80$ to $Y = -40$).

On the other hand, inattention eradicated the differences between sentences and word lists. The interaction of attended > unattended X sentences > word lists was significant

from the inferior temporal pole (MNI -35 -5 -36, $Z = 4.88$) to the anterior fusiform gyrus (MNI -40 -39 -23, $Z = 3.73$) (corrected within the left VOT region of interest; Fig. 4, bottom row). When restricted to unattended written stimuli, the sentences > lists contrast did not activate a single voxel.

3.5. Activation by spoken language

3.5.1. Attended spoken language

Relative to baseline, the VOT cortex was activated by spoken language and deactivated by control stimuli, anterior to about $Y = -60$ (Fig. 4, top row). The spoken language > control contrast activated the left VOT cortex from the inferior temporal pole (MNI -40 -7 -41, $Z = 5.74$), including the supramodal region around $Y = -35$ described before (MNI -45 -32 -18, $Z = 6.84$), and extending caudally back to about $Y = -60$ (MNI -50 -54 -18,

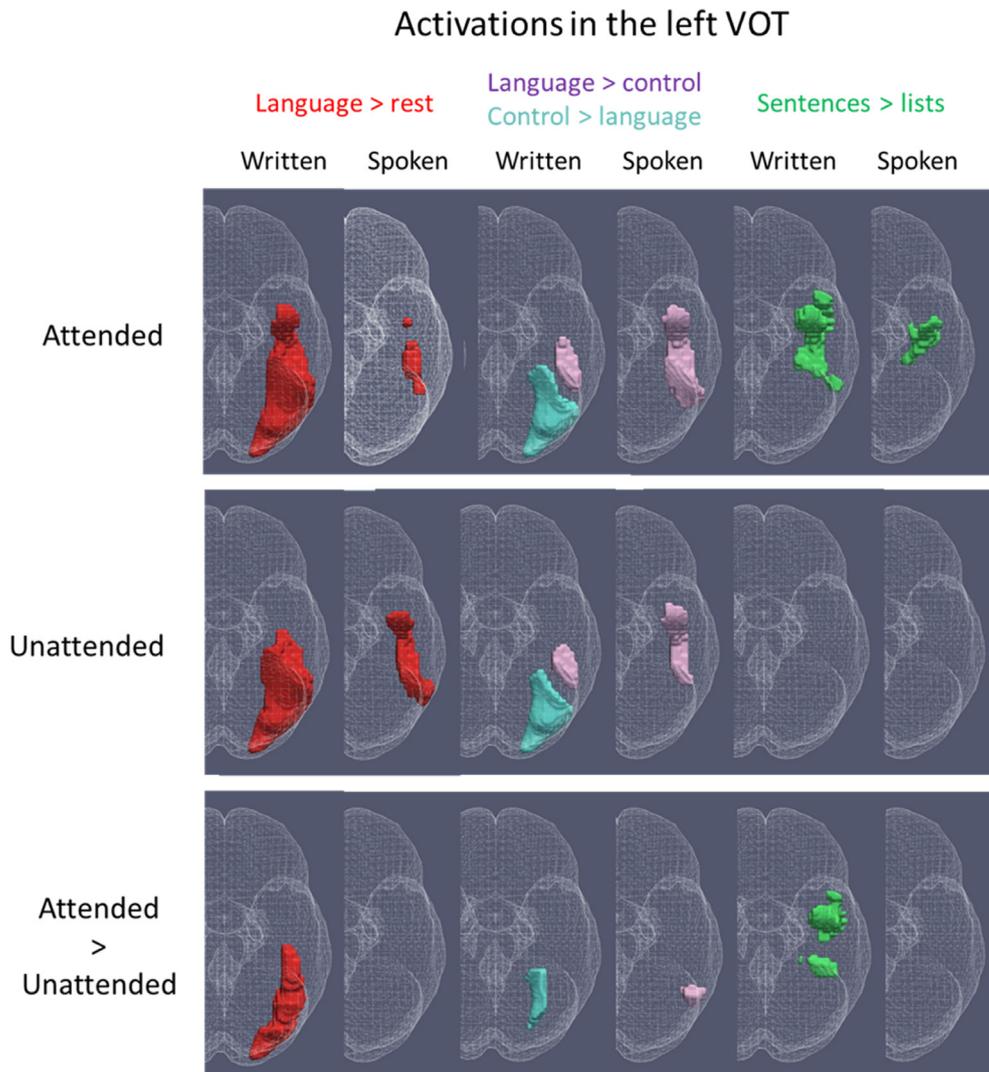


Fig. 4 – Ventral view of activations restricted to the left ventral occipito-temporal region of interest. Activations are displayed on a transparent brain, separately for written and spoken stimuli (odd and even columns, respectively), for attended (top row) and unattended (middle row) stimuli, and their difference (bottom row). In each row are represented activations by language (averaging sentences and lists) > rest (red), control > language (light blue), language > control (pink), and sentences > word lists (green).

$Z = 4.83$), close to the usual location of the VWFA. When contrasting language minus baseline, activations were smaller but still extended back to the VWFA (MNI -50 -52 -21, $Z = 3.27$).

3.5.2. Attended spoken sentences

The contrast of spoken sentences minus lists showed VOT activations in the inferior temporal pole (MNI -42 -10 -38, $Z = 3.51$) and anterior fusiform gyrus (Figs. 4D and 5G; MNI -40 -19 -26, $Z = 4.47$).

3.5.3. Effect of attention on spoken language processing

Removing attention away from speech stimuli did not significantly reduce the activation by language relative to baseline, while it suppressed the deactivation by spoken control. This is why inattention reduced the activation by spoken language > control in the vicinity of the VWFA (MNI -52 -59 -18, $Z = 3.69$, corrected within the left VOT ROI; Fig. 4, bottom row).

In the unattended condition, the contrast of language > control remained significant in an extended VOT cluster ranging from the inferior temporal pole to the fusiform region, back to $Y = -49$ (Fig. 4, middle row). The same was even more true when contrasting language minus baseline (MNI -47 -49 -23, $Z = 5.10$).

Under inattention, there was no residual activation when contrasting sentences > lists, although the reduction of this contrast by attention removal did not reach significance.

4. Discussion

4.1. Whole-brain supramodal activations

Whole-brain analyses showed supramodal activations by attended language minus controls in left frontal and temporal

semantic processes. Indeed, in a meta-analysis, Binder et al. (2009) locate a semantic focus at MNI $Y = -35$, that is precisely where peak the current activations.

Still at the whole-brain level, attended sentences minus lists activated a subset of the language network, including the core syntax areas in the STS and IFG (Pallier et al., 2011; Snijders et al., 2009; Tyler et al., 2011), plus two anterior temporal lobe (ATL) regions, namely the superior temporal pole and the most anterior fusiform cortex ($Y = -19$). Similarly, in a sample of over 200 participants, Uddén et al. (2019) compared various types of sentences and word lists, finding no modality-specific effects related to sentence unification processes (see also Braze et al., 2011; Constable et al., 2004; Michael et al., 2001a). One may note that those studies resorted only to unimodal stimulation. In contrast we used an attentionally demanding multimodal set-up, possibly making written stimuli somewhat easier to perceive than spoken stimuli (see Supplementary Figure 1). This may explain why the network was activated more strongly by sentence-level processing with written than with spoken stimuli (Fig. 2B).

Left ATL activations have been repeatedly reported in relation to the processing of sentences as compared to word lists or similar controls, and have been proposed to reflect combinatorial semantics rather than syntax per se (for reviews see Pallier et al., 2011; Pyllkänen, 2016; Rogalsky, 2016), although recent studies cast doubt on this notion (Fedorenko et al., 2016). Note that ventral ATL activations are found less systematically than superior polar ones, as they fall in a region of susceptibility-related signal loss in the MNI $Y = -10$ to -30 range (Binder et al., 2011). Here we successfully addressed this technical issue by using a triple-echo imaging sequence (Poser et al., 2006).

In agreement with early demonstrations that only reduced information is extracted from unattended speech (Broadbent, 1962; Cherry, 1953; Treisman, 1964), inattention (here caused by drawing attention away from the modality in which language was presented) led to a drastic reduction in most activations by language minus controls, and to the absence of any significant activations by sentences minus lists of words, despite a residual trend in favor of sentences in Bayesian analyses (Supplementary Figure 3). This supports the view that significant single word processing persists under inattention, but that sentence-level processing is largely disrupted, if not entirely cancelled (Batterink & Neville, 2013; Gui et al., 2020; Nakamura, Makuuchi, Oga, Mizuochi-Endo, et al., 2018).

In the present case, it is likely that residual word-level processing was possible because, in our paradigm, unattended stimuli were physically intact and identical to attended stimuli. Indeed, when words are made unconscious through perceptual degradation or masking, their processing then requires attentional support. The processing of masked words has been demonstrated behaviorally up to the semantic level using priming methods, both with written and spoken stimuli (Kouider & Dehaene, 2007; Kouider & Dupoux, 2005), with corresponding word-related brain activations (Dehaene et al., 2001; Kouider et al., 2010; Kutas & Federmeier, 2011). However, such priming effects disappear whenever attention is focused away from the time window (Kiefer & Brendel, 2006; Naccache et al., 2002) or away from the location (Lachter et al.,

2004) where masked primes are presented (for reviews see Finkbeiner & Forster, 2008; Kouider & Dehaene, 2007).

In contrast, sentence-level processing seems to require both attention and unmasked stimulation. When attention is actively diverted away from intact language stimuli, ERPs show sensitivity to violations of local word agreement (Maidhof & Koelsch, 2011; Pulvermuller et al., 2008) but not to violations of higher-level syntactic violations (Lee et al., 2019). When, conversely, words are masked although attended, there is also evidence of integration of contiguous words (e.g., Batterink & Neville, 2013; Nakamura, Makuuchi, Oga, Mizuochi-Endo, et al., 2018; van Gaal et al., 2014), while long-distance syntactic integration does not receive consistent support. For instance, the incongruence between short sentences and target pictures elicited a N400 signature only whenever sentences were consciously perceived, while a N400 wave occurred even for masked single words (Mongelli et al., 2019).

In summary, our results converge with others in supporting the idea that multi-word language processing, particularly the integration of words into syntactically and semantically correct phrases, requires attention and consciousness, for both input modalities.

4.2. Left VOT activations

Analyses focused on the left VOT showed that this area could be activated not only bottom-up by written words but also top-down by spoken language, and that it was modulated by further top-down parameters including attention orientation and sentence-level processing. The pattern of activation in the left VOT may be summarized by distinguishing 3 sectors, from the back to the front of the VOT cortex (Fig. 5).

4.3. The postero-mesial VOT cortex

This is the area showing stronger bottom-up activation for control stimuli (alphabetical stimuli in an unknown alphabet), than for real words (Fig. 4, first row, third column, light blue). Here, changes in attention had a significant impact, as inattention for visual input decreased activation level for all visual stimuli (Kanwisher & Wojciulik, 2000). Nevertheless, inattention had a larger effect on control stimuli than on real words, suggesting that despite their lower absolute level, activations by printed words were more robust to attentional diversion than unfamiliar control stimuli. This observation fits with the evidence reviewed before of high-level processing of unattended or even unconscious words. Importantly, however, this region showed no effect of the other two top-down factors, as we found no activation by spoken language, and no difference between sentences and lists of words. Thus, this region activates in a purely bottom-up visual manner and presumably carries out processing of an early visual nature.

4.4. The mid-lateral VOT cortex and the VWFA

This is the area showing stronger bottom-up activation for printed words than for control stimuli (Fig. 4, first row, third column, pink color). This area shows the opposite pattern of visual activation as compared to the postero-mesial VOT

cortex: Printed words yielded higher activation than the unfamiliar alphabet, which did not activate this area above baseline. The oblique line of transition between the posteromesial and the mid-lateral regions is a clear illustration of language specialization in more anterior sectors of the VOT cortex. The superiority of the activation by printed words was not significantly altered by inattention, again pointing to the automaticity of reading processes.

The VWFA proper showed ambiguous results, as its usual peak coordinates place it at the boundary between those two zones, where the difference between written words and controls is smallest ($p > .05$ at MNI = $-42 -57 -15$; Cohen et al., 2002). Unfortunately, our design did not allow us to positively identify the VWFA. This is likely due to the fact that we used as control an unknown alphabet visually similar to the familiar alphabet, rather than more contrastive baselines such as other categories of images (e.g., Baeck et al., 2015; Monzalvo et al., 2012) or low-level stimuli such as checkers (Cohen et al., 2000, 2002; Mano et al., 2012). Although a superiority of familiar over unfamiliar alphabets in the VWFA was reported in some studies (e.g., Baker et al., 2007), it was absent in others (e.g., Vinckier et al., 2007), likely depending on the way the VWFA region of interest was defined, as well as on task demands.

Importantly, the mid-lateral VOT cortex showed richer top-down influences than the posterior VOT region. Thus, it was strongly activated in a top-down manner by speech relative to control noise. Activation by speech in the left VOT cortex was found previously using single spoken words (MNI $-45 -64 -11$; Ludersdorfer et al., 2016) or sentences (MNI $-42 -43 -15$; Planton et al., 2019).

We found that activations by speech at in the vicinity of the VWFA (Cohen et al., 2000, 2002), was more consistent when subtracting the control noise condition, than relative to rest. This observation matches the pattern reported about the VWFA by Yoncheva et al. (2010) and Ludersdorfer et al. (2016), showing that (1) various auditory stimuli induce a general deactivation of visual cortex relative to rest; and (2) within this broad deactivation, the VWFA behaves as an “island” of lesser or no deactivation by speech stimuli. We and others found that this speech versus noise difference is critically dependent on attention: across various paradigms, it emerges or is reinforced when speech is attended (Fig. 4, third row), when attention is oriented towards linguistic features of stimuli (Yoncheva et al., 2010), and during a demanding auditory lexical decision task (as used here), relative to the mere listening to spoken sentences (Dehaene et al., 2011).

There are several indications that, in literate participants at least, this activation of the VWFA by speech reflects the computation of orthographic representations from auditory input. First, it is strongly modulated by literacy, and close to nil in illiterate subjects (Dehaene et al., 2011). Second the activations observed by Ludersdorfer et al. (2016) were present during an orthographic but not a semantic task, supporting top-down activation of orthographic codes. Furthermore Planton et al. (2019) showed that there was substantial spatial overlap, and significant correlation across voxels, between word reading and speech perception.

Finally, this region showed stronger activation to sentences than lists, but only for attended written input. As

discussed below, a much stronger difference between sentences and word lists was found more anteriorly in inferior temporal cortex, as well as throughout the left STS and inferior frontal gyrus. Thus, it seems likely that the sentence effects reflect an effortful top-down maintenance of orthographic information on the incoming written words when participants have to integrate them into phrases. While fMRI lacks temporal resolution, this proposal could be tested with a time-resolved method such as MEG or intracranial recordings: in mid-temporal cortex, an early bottom-up activation would be followed by a sustained activity only if the words can be integrated in a sentence. Such an effect may have been missed in existing studies, either because they lacked ventral temporal electrodes or because they focused only on bottom-up transient effects (Fedorenko et al., 2016; Nelson et al., 2017; Woolnough et al., 2020).

One may summarize this region’s activation to speech, as follows: (1) The VWFA activates during speech perception relative to background deactivation by auditory stimuli, (2) probably reflecting a top-down access to orthographic codes, (3) mostly in the presence of attention and as required by task demands.

4.5. The anterior VOT cortex

Anterior to about $Y = -30$, we found activations for both written language (minus rest) and spoken language (minus control). Those activations were almost double in size for full sentences compared to word lists (Fig. 5), and this effect of sentence structure prevailed in both modalities. Under inattention, activations weakened and the sentence-integration effect vanished entirely. As discussed with whole-brain activations, this pattern identifies the ventral ATL as a supramodal and integrative language area. It is congruent with a considerable amount of prior research on the involvement of this area in syntactic and particularly semantic integration, not only for spoken and written sentences (for reviews see Pallier et al., 2011; Pykkänen, 2016; Rogalsky, 2016), but also whenever multiple concepts need to be combined into a coherent role, such as in the Pyramid-and-Palm-Tree test with pictures (Vandenberghe et al., 1996).

5. Conclusion

We started this paper with two questions. First, can sentence-level processing occur for unattended stimuli? The answer to this question seems to be largely negative: the withdrawal of attention drastically reduces the difference between sentences and word lists, down to an undetectable level. This does not mean that, with more sensitive methods, subtle residual processing could not be detected, but it does indicate that inattention drastically affects sentence-level integration, perhaps only down to local interactions between nearby words (Batterink & Neville, 2013; Gui et al., 2020; Nakamura, Makuuchi, Oga, Mizuochi-Endo, et al., 2018).

Second, we asked whether inattention would cancel all top-down effects on VOT cortex, and if so, whether it would also cancel the well-known specialization of part of the VOT (particularly the VWFA) for orthographically readable stimuli,

thus supporting the view (Price and Devlin, 2003, 2011) that there is no genuine specialization for reading in this area, but only a top-down task-related recruitment. Here, the results must be carefully unpacked. On the one hand, within written stimuli, we clearly found that the differential response of VOT to known versus unknown alphabets did not vanish under inattention – it was reduced, but the stronger posterior response to the unknown alphabet and, most importantly, the stronger mid-lateral VOT response to the known alphabet remained. Since this language>control activation was reduced to this site and vanished from elsewhere in the brain in the unattended condition (Fig. 3B, right panel), it seems unlikely that it could have arisen from a top-down effect, contra Price and Devlin, (2003, 2011). Rather, it must therefore have arisen in a purely bottom-up manner, as a reflection of the specialized and automatic processing of letter strings at this site (Dehaene & Cohen, 2011; Dehaene et al., 2005).

On the other hand, with spoken language, we found a clear top-down effect, which surprisingly remained present for unattended stimuli: spoken language yielded a greater activation than auditory controls in mid-lateral and anterior VOT, with a small difference between attended and unattended conditions (Fig. 5). Furthermore, in anterior VOT, another type of effect, possibly of top-down origin, was the superiority of sentences over word lists. The conclusion which arises from those findings is that top-down effects of auditory stimuli on the ventral visual pathway are genuine, including in the VWFA (as previously reported: Dehaene et al., 2011; Planton et al., 2019; Yoncheva et al., 2010), and that the methodology of withdrawing attention may not be satisfactory if the goal is to prevent such top-down effects. The VWFA is likely to be traversed by multiple language-related signals, some related to early bottom-up processing and others to a later wave of top-down influences (Thesen et al., 2012; Whaley et al., 2016; Woodhead et al., 2014; Woolnough et al., 2020).

Those conclusions must be qualified by the limits of the present study, which is a group study with relatively limited spatial resolution and no distinct subject-specific localizer for the VWFA proper. Still a further caveat should be kept in mind concerning the efficacy of the method we used to divert attention away from language. High performance with attended non-language stimuli showed that attention was sufficiently oriented towards this stream to allow for its successful processing. However, what it implies for the processing of the simultaneous unattended language stream is debatable. On the one hand, if performance had been poor, the most straightforward interpretation would have been that subjects were distracted away from non-language stimuli towards language processing. In this view, good performance suggests that the design achieved the desired orientation of attention away from language. On the other hand, high performance on the non-language task indicates that abundant attentional resources were available. The surplus of attentional capacity not taken up by the task may then have spilled over to task-irrelevant language stimuli. In the context of the load theory of attention, this would be a case of a task with low perceptual load, allowing for the task-irrelevant stimuli to be processed beyond the perceptual level (Lavie, 2005; Lavie & Dalton, 2014, pp. 56–75).

In summary, we have no direct behavioral measure of the level of processing of unattended language. Poor performance on attended non-language stimuli would have been a matter of concern, suggesting some irrepressible orientation of attention towards language. The good performance which prevailed, however, is not sufficient proof that attention was entirely removed from unattended language. Imaging results, however, showing strongly reduced activation under inattention, suggest that our attention manipulation away from language was effective, if perhaps not complete.

The present work paves the way to a more detailed study contrasting words with a variety of other categories of stimuli such as numbers or faces, using finer-grained fMRI images, single-subject analyses, and possibly laminar resolution (Sharoh et al., 2019), such that bottom-up and top-down influences could be ultimately distinguished by their layer-specific signatures.

CrediT author statement

Conceptualization: LC, PS, CP, SD; Data curation: PS, CP; Formal analysis: LC, PS; Funding acquisition LC; Project administration: LC; Writing - review & editing: LC, SD.

Open practices

The study in this article earned Open Materials and Open Data badges for transparent practices.

Acknowledgments

This work has received funding from the program “Investissements d’avenir” ANR-10-IAIHU-06, and was supported by INSERM, CEA, and Collège de France. We are grateful to Benoît Béranger at the CENIR.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2021.02.007>.

REFERENCES

- Baeck, A., Kravitz, D., Baker, C., & Op de Beeck, H. P. (2015). Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area. *Neuroimage*, 111, 321–328.
- Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences U S A*, 104, 9087–9092.
- Barttfeld, P., Abboud, S., Lagercrantz, H., Adén, U., Padilla, N., Edwards, A., Cohen, L., Sigman, M., Dehaene, S., & Dehaene-

- Lambertz, G. (2018). A lateral-to-mesial organization of human ventral visual cortex at birth. *Brain Structure & Function*, 1–13.
- Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, 33, 8528–8533.
- Bemis, D. K., & Pyllkanen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23, 1859–1873.
- Benjamin, C. F., Walshaw, P. D., Hale, K., Gaillard, W. D., Baxter, L. C., Berl, M. M., Polczynska, M., Noble, S., Alkawadri, R., Hirsch, L. J., Constable, R. T., & Bookheimer, S. Y. (2017). Presurgical language fMRI: Mapping of six critical regions. *Human Brain Mapping*, 38, 4239–4255.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Science*, 15, 527–536.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19, 2767–2796.
- Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., Grabowski, T. J., Langfitt, J. T., Loring, D. W., Lowe, M. J., Koenig, K., Morgan, P. S., Ojemann, J. G., Rorden, C., Szaflarski, J. P., Tivarus, M. E., & Weaver, K. E. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *Neuroimage*, 54, 1465–1475.
- Bouhali, F., Bezagu, Z., Dehaene, S., & Cohen, L. (2019). A mesial-to-lateral dissociation for orthographic processing in the visual cortex. *Proceedings of the National Academy of Sciences U S A*, 116, 21936–21946.
- Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J. F., Dehaene, S., & Cohen, L. (2014). Anatomical connections of the visual word form area. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34, 15402–15414.
- Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Constable, R. T., Fulbright, R. K., Magnuson, J. S., Van Dyke, J. A., & Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, 47, 416–431.
- Broadbent, D. E. (1962). Attention and the perception of speech. *Scientific American*, 3.
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, Psychophysics*, 77, 1465–1487.
- Buchweitz, A., Mason, R. A., Tomitch, L. M., & Just, M. A. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology and Neuroscience*, 2, 111–123.
- Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18, 90–98.
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25, 975–979.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123, 291–307.
- Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage*, 40, 353–366.
- 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.
- Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *Neuroimage*, 22, 11–21.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers Biomedical Research*, 29, 162–173.
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. J. P.b. (2018). *The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition* (Vol. 16), Article e2004103.
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Science*, 15, 254–262.
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16, 234–244.
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Science*, 9, 335–341.
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4, 752.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330, 1359–1364.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes Filho, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2011). Impact on the brain of learning to read. *Medical Science (Paris)*, 27, 236–238.
- Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., Umiltà, C., & Mascetti, G. G. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology*, 23, 841–855.
- Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, 113, E6256–E6262.
- Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, 18, 120–126.
- Finkbeiner, M., & Forster, K. I. (2008). Attention, intention and domain-specific processing. *Trends in Cognitive Sciences*, 12, 59–64.
- Forte, A. E., Etard, O., & Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife*, 6, Article e27203.
- Friston, K. J., Worsley, K. J., Frackowiak, R. S., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1, 210–220.
- Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's area. *Neuroimage*, 114, 294–302.
- Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A vision of reading. *Trends in Cognitive Sciences*, 20, 171–179.
- Gui, P., Jiang, Y., Zang, D., Qi, Z., Tan, J., Tanigawa, H., Jiang, J., Wen, Y., Xu, L., & Zhao, J. (2020). Assessing the depth of language processing in patients with disorders of consciousness. *Nature Neuroscience*, 1–10.
- Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends in Cognitive Sciences*, 19, 374–382.

- Hill, K. T., & Miller, L. M. (2009). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, 20, 583–590.
- Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., & Orban, G. A. (2012). Processing of vocalizations in humans and monkeys: A comparative fMRI study. *Neuroimage*, 62, 1376–1389.
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews. Neuroscience*, 1, 91–100.
- Kiefer, M., & Brendel, D. (2006). Attentional modulation of unconscious “automatic” processes: Evidence from event-related potentials in a masked priming paradigm. *Journal of Cognitive Neuroscience*, 18, 184–198.
- 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., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science*, 16, 617–625.
- Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*, 60(3), 1759–1770.
- Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vértes, P. E., Inati, S. J., ... Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences*, 110(40), 16187–16192.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review*, 111, 880.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends Cognitive Science*, 9, 75–82.
- Lavie, N., & Dalton, P. (2014). Load theory of attention and cognitive control. In *The Oxford handbook of attention*. Oxford: Oxford University Press.
- Lee, D. J., Jung, H., & Loui, P. (2019). Attention modulates electrophysiological responses to simultaneous music and language syntax processing. *Brain sciences*, 9.
- Li, J., Osher, D. E., Hansen, H. A., & Saygin, Z. M. (2019). Cortical selectivity driven by connectivity: Innate connectivity patterns of the visual word form area. *bioRxiv*, 712455.
- Lobier, M., Zoubrinetzky, R., & Valdois, S. (2012). The visual attention span deficit in dyslexia is visual and not verbal. *Cortex*, 48, 768–773.
- Ludersdorfer, P., Wimmer, H., Richlan, F., Schurz, M., Hutzler, F., & Kronbichler, M. (2016). Left ventral occipitotemporal activation during orthographic and semantic processing of auditory words. *Neuroimage*, 124, 834–842.
- Lüders, H., Lesser, R. P., Hahn, J., Dinner, D. S., Morris, H. H., Wylie, E., & Godoy, J. (1991). Basal temporal language area. *Brain*, 114, 743–754.
- Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience*, 23, 2252–2267.
- Mano, Q. R., Humphries, C., Desai, R. H., Seidenberg, M. S., Osmon, D. C., Stengel, B. C., & Binder, J. R. (2012). The role of left occipitotemporal cortex in reading: Reconciling stimulus, task, and lexicality effects. *Cerebral Cortex*, 23(4), 988–1001.
- Mazoyer, B. M., Dehaene, S., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Cohen, L., & Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5, 467–479.
- McDermott, J. H. (2009). The cocktail party problem. *Current Biology*, 19, R1024–R1027.
- Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485, 233.
- Michael, E. B., Keller, T. A., Carpenter, P. A., & Just, M. A. (2001a). fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. *Human Brain Mapping*, 13, 239–252.
- Michael, E. B., Keller, T. A., Carpenter, P. A., & Just, M. A. (2001b). fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, 13, 239–252.
- Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., Fryer, T. D., Williams, G. B., Hodges, J. R., & Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain*, 133, 3256–3268.
- Mongelli, V., Meijs, E. L., van Gaal, S., & Hagoort, P. (2019). No language unification without neural feedback: How awareness affects sentence processing. *Neuroimage*, 202, 116063.
- Monzalvo, K., Fluss, J., Billard, C., Dehaene, S., & Dehaene-Lambertz, G. (2012). Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status. *Neuroimage*, 61, 258–274.
- Moreno, A., Limousin, F., Dehaene, S., & Pallier, C. (2018). Brain correlates of constituent structure in sign language comprehension. *Neuroimage*, 167, 151–161.
- Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science*, 13, 416–424.
- Nakamura, K., Makuuchi, M., Oga, T., Mizuuchi-Endo, T., Iwabuchi, T., Nakajima, Y., & Dehaene, S. (2018a). Neural capacity limits during unconscious semantic processing. *The European Journal of Neuroscience*, 47, 929–937.
- Nakamura, K., Makuuchi, M., Oga, T., Mizuuchi-Endo, T., Iwabuchi, T., Nakajima, Y., & Dehaene, S. (2018b). Neural capacity limits during unconscious semantic processing. *European Journal of Neuroscience*, 47, 929–937.
- Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S. S., Naccache, L., Hale, J. T., Pallier, C., & Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114, E3669–E3678.
- Nichols, T. E. (2012). Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage*, 62, 811–815.
- Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108, 2522–2527.
- Pattamadilok, C., Chanoine, V., Pallier, C., Anton, J.-L., Nazarian, B., Belin, P., & Ziegler, J. C. (2017). Automaticity of phonological and semantic processing during visual word recognition. *Neuroimage*, 149, 244–255.
- Pinel, P., Thirion, B., Meriaux, S., Jobert, A., Serres, J., Le Bihan, D., Poline, J. B., & Dehaene, S. (2007). Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neuroscience*, 8, 91–109.
- Planton, S., Chanoine, V., Sein, J., Anton, J. L., Nazarian, B., Pallier, C., & Pattamadilok, C. (2019). Top-down activation of the visuo-orthographic system during spoken sentence processing. *Neuroimage*, 202, 116135.
- Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, 55, 1227–1235.

- Price, C. J. (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, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage*, 19, 473–481.
- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15, 246–253.
- Pulvermuller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language*, 104, 244–253.
- Pylkkänen, L. (2016). Composition of complex meaning: Interdisciplinary perspectives on the left anterior temporal lobe. In *Neurobiology of language* (pp. 621–631). Elsevier.
- Rogalsky, C. (2016). The role of the anterior temporal lobe in sentence processing. In *Neurobiology of language* (pp. 587–595). Elsevier.
- Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitel, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage*, 39, 1444–1456.
- Saygin, Z. M., Osher, D. E., Koldewyn, K., Reynolds, G., Gabrieli, J. D., & Saxe, R. R. (2012). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature Neuroscience*, 15, 321–327.
- Sharoh, D., van Mourik, T., Bains, L. J., Segaert, K., Weber, K., Hagoort, P., & Norris, D. G. (2019). Laminar specific fMRI reveals directed interactions in distributed networks during language processing. *bioRxiv*, 585844.
- Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An FMRI study using word-category ambiguity. *Cerebral Cortex*, 19, 1493–1503.
- Taylor, J. S. H., Davis, M. H., & Rastle, K. (2019). Mapping visual symbols onto spoken language along the ventral visual stream. *Proceedings of the National Academy of Sciences*, 116(36), 17723–17728.
- Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., & Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications*, 3, 1284.
- Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning Verbal Behavior*, 3, 449–459.
- Tyler, L. K., Marslen-Wilson, W. D., Randall, B., Wright, P., Devereux, B. J., Zhuang, J., Papoutsis, M., & Stamatakis, E. A. (2011). Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain*, 134, 415–431.
- Uddén, J., Hultén, A., Schöffelen, J.-M., Lam, N., Harbusch, K., van den Bosch, A., Kempen, G., Petersson, K. M., & Hagoort, P. (2019). Supramodal sentence processing in the human brain: fMRI evidence for the influence of syntactic complexity in more than 200 participants. *bioRxiv*, 576769.
- van Gaal, S., Naccache, L., Meuwese, J. D., van Loon, A. M., Leighton, A. H., Cohen, L., & Dehaene, S. (2014). Can the meaning of multiple words be integrated unconsciously? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369, 20130212.
- Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383, 254–256.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron*, 55, 143–156.
- Whaley, M. L., Kadipasaoglu, C. M., Cox, S. J., & Tandon, N. (2016). Modulation of orthographic decoding by frontal cortex. *Journal of Neuroscience*, 36, 1173–1184.
- Woodhead, Z. V., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., & Leff, A. P. (2014). Reading front to back: MEG evidence for early feedback effects during word recognition. *Cerebral Cortex*, 24, 817–825.
- Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2020). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *bioRxiv*.
- Yoncheva, Y. N., Zevin, J. D., Maurer, U., & McCandliss, B. D. (2010). Auditory selective attention to speech modulates activity in the visual word form area. *Cerebral Cortex*, 20, 622–632.