Top-down activation of the visuo-orthographic system during spoken sentence processing

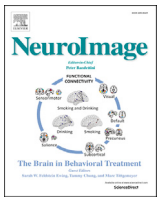
Samuel Planton, Valérie C Chanoine, Julien Sein, Jean-Luc Anton, Bruno Nazarian, Christophe C Pallier, Chotiga Pattamadilok

NeuroImage 202 (2019) 116135

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Top-down activation of the visuo-orthographic system during spoken sentence processing

Samuel Planton [a,d,*], Valérie Chanoine [b], Julien Sein [c], Jean-Luc Anton [c], Bruno Nazarian [c], Christophe Pallier [d], Chotiga Pattamadilok [a]

[a] Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[b] Aix Marseille Univ, Institute of Language, Communication and the Brain, Brain and Language Research Institute, Aix-en-Provence, France
[c] Aix Marseille Univ, CNRS, Centre IRM-INT, INT UMR, 7289, Marseille, France
[d] INSERM-CEA, Cognitive Neuroimaging Unit, Neurospin Center, Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
fMRI
Left ventral occipitotemporal cortex
Natural speech processing
Speech in noise
Visual word form area

ABSTRACT

The left ventral occipitotemporal cortex (vOT) is considered the key area of the visuo-orthographic system. However, some studies reported that the area is also involved in speech processing tasks, especially those that require activation of orthographic knowledge. These findings suggest the existence of a top-down activation mechanism allowing such cross-modal activation. Yet, little is known about the involvement of the vOT in more natural speech processing situations like spoken sentence processing. Here, we addressed this issue in a functional Magnetic Resonance Imaging (fMRI) study while manipulating the impacts of two factors, i.e., task demands (semantic vs. low-level perceptual task) and the quality of speech signals (sentences presented against clear vs. noisy background). Analyses were performed at the levels of whole brain and region-of-interest (ROI) focusing on the vOT voxels individually identified through a reading task. Whole brain analysis showed that processing spoken sentences induced activity in a large network including the regions typically involved in phonological, articulatory, semantic and orthographic processing. ROI analysis further specified that a significant part of the vOT voxels that responded to written words also responded to spoken sentences, thus, suggesting that the same area within the left occipitotemporal pathway contributes to both reading and speech processing. Interestingly, both analyses provided converging evidence that vOT responses to speech were sensitive to both task demands and quality of speech signals: Compared to the low-level perceptual task, activity of the area increased when efforts on comprehension were required. The impact of background noise depended on task demands. It led to a decrease of vOT activity in the semantic task but not in the low-level perceptual task. Our results provide new insights into the function of this key area of the reading network, notably by showing that its speech-induced top-down activation also generalizes to ecological speech processing situations.

1. Introduction

Studies of the neural basis of language processing have revealed the involvement of a wide range of brain regions distributed over the temporal, frontal and occipital lobes (Price, 2012). These widely distributed networks reflect the contribution of both modality-specific sensory-motor and shared higher-level cognitive systems during spoken and written language processing. Spoken language processing is carried out through a succession of stages from acoustic signal processing in bilateral auditory cortices to phonological processing and lexical access in left-dominant posterior and middle superior temporal sulcus (STS) (Hickok and Poeppel, 2007). Visual word processing, on the other hand, recruits a hierarchically organized occipitotemporal pathway in the left hemisphere, allowing access to orthographic information form visual input (Dehaene et al., 2005; Vinckier et al., 2007). Within this "reading pathway", the ventral occipitotemporal cortex (vOT) in the left mid-fusiform gyrus, also known as the "Visual Word Form Area" (VWFA), plays a key role as demonstrated by its sensitivity to visual orthographic input (Cohen et al., 2000; Dehaene and Cohen, 2011; McCandliss et al., 2003; Turkeltaub et al., 2002). Its location is remarkably reproducible across individuals, cultures and writing systems (Bolger et al., 2005), and it has been shown to become increasingly specialized with reading

* Corresponding author. F/JOLIOT/NEUROSPIN/UNICOG, Bât. 145 - Point Courrier 156, F-91191, Gif sur Yvette Cedex, France.
E-mail address: samuel.planton@cea.fr (S. Planton).

https://doi.org/10.1016/j.neuroimage.2019.116135
Received 16 June 2019; Received in revised form 9 August 2019; Accepted 26 August 2019
Available online 27 August 2019

experience (Brem et al., 2010; Dehaene-Lambertz et al., 2018).

Despite its well-established role in reading and its location in the extrastriate visual cortex, the specificity of left vOT responses to visual inputs has been challenged on multiple occasions (Büchel et al., 1998; Ludersdorfer et al., 2016; Price and Devlin, 2003; Reich et al., 2011). Several brain-imaging studies have shown that this area can be activated by speech, in the absence of any visual sensory input. This is especially the case when the tasks require explicit access to orthographic representations. For instance, when participants have to judge whether spoken words share the same rime spelling (Booth et al., 2002; Cao et al., 2010), contain a target letter (Ludersdorfer et al., 2015) or contain a specific number of letters (Ludersdorfer et al., 2016). Furthermore, a correlation between the activity within this area and the performance in an auditory spelling task was reported (Booth et al., 2003). Such cross-modal activations have mainly been explained by a task-induced top-down activation of orthographic representations stored in the visuo-orthographic processing pathway (Ludersdorfer et al., 2015, 2016).

Interestingly, vOT activations have also been reported in speech recognition tasks that do not require the explicit recovery of orthographic information, although the findings are less robust. Yoncheva et al. (2010) presented participants with pairs of spoken words overlaid with tone triplets. They found that the vOT was more strongly activated when the participants selectively attended to words' rimes than to tone-triplets. In line with this observation, Ludersdorfer et al. (2016), reported that the amplitude of vOT responses to speech depended on task demands, i.e., were highest and most wide-spread in a spelling task, significantly reduced in a semantic task and absent in an acoustic task. Taken together, the existing literature suggests that, in literate individuals, the top-down vOT activation occurs whenever the linguistic content of speech is processed. However, the degree of activation seems to vary with the contexts in which speech is processed.

So far, most studies that investigated vOT responses to speech have been conducted on single words or pseudowords and have been using relatively artificial speech processing situations (Booth et al., 2002, 2004; Burton et al., 2005; Cohen et al., 2004; Cone et al., 2008; Desroches et al., 2010; Seidenberg and Tanenhaus, 1979). Thus, the question remains whether the cross-modal activation of this area also generalizes to more ecological speech processing situations. One example of these situations is when participants are required to process spoken sentences. To our knowledge, the only relevant finding was reported by Dehaene et al. (2010) who examined brain activity in illiterate and literate participants during passive sentence listening and an auditory lexical decision task. While the activity of the left vOT observed in the auditory lexical decision task increased with participants' reading experience, it remained absent during passive sentence listening in both populations. This latter observation strongly questions the idea of automatic bidirectional coupling between spoken and written language (Grainger and Ziegler, 2007; Harm and Seidenberg, 1999, 2004). Rather, it suggests that the recruitment of the orthographic system during speech processing is optional and maybe restricted to artificial experimental setups (Dehaene and Cohen, 2011). In addition to the role of task demands, the null result reported by Dehaene et al. (2010) could also be related to the fact that, unlike single word processing, the conversion from phonology to orthography at the sentence level is far too elaborated and therefore unlikely, even in the most challenging speech processing contexts.

In addition to the theoretical debate on the automaticity of vOT activation in response to spoken input, several pieces of evidence also pointed to a methodological issue related to how the activation of the vOT has been computed. Indeed, the "significant" vOT activation reported in certain studies did not necessarily reflect an increase of vOT activation compared to a fixation (silent) baseline but rather a "reduced deactivation" of this area compared to when participants processed non-linguistic auditory stimuli (Ludersdorfer et al., 2016; Yoncheva et al., 2010). A reduction of activity in the visual system during auditory input processing has indeed been reported and explained by a cross-modal sensory suppression phenomenon (Laurienti et al., 2002). In line with this observation, the few studies that showed significant vOT activation in non-orthographic or low-level speech processing tasks (e.g., one-back task) were those that contrasted brain activity observed in speech processing conditions to that observed in non-speech auditory conditions (e.g., auditory tones, time-reversed speech). Since the activity of the visual system was strongly suppressed in the latter conditions (Ludersdorfer et al., 2013, 2016), it remains unclear to what extent the seemingly "significant" vOT activation truly reflected the contribution of this area in speech processing.

Another methodological issue relates to the specific localization of the area within the ventral part of the left occipitotemporal pathway that is involved in speech processing, more particularly, whether it corresponds precisely to the area commonly referred to as the VWFA. In the original work describing the VWFA, Cohen et al. (2004) also reported a region, namely the lateral inferotemporal multimodal area (LIMA) that was activated by both spoken and written words. This area is also located in the ventral occipitotemporal pathway, lateral and anterior to the VWFA. Using the regions-of-interest (ROIs) whose coordinates correspond to the activation peaks reported in the visual-word processing literature is an approach that allows us to address this issue. As different portions of the left fusiform gyrus show different degrees of selectivity to written stimuli (posterior-anterior gradient of activation; Vinckier et al., 2007), some authors chose to build series of ROIs along the ventral visual pathway (Ludersdorfer et al., 2013, 2015, 2016; Yoncheva et al., 2010). The finding that the activity induced by spoken inputs is located in the portion of the pathway that responds to written inputs has been used as an argument in favor of a homology between the area activated by spoken words and the VWFA described in the literature.

However, even though individual brains are generally aligned to a common template brain in the same stereotaxic space, comparisons of brain activity obtained in different studies based on Talairach or MNI coordinates are subject to many sources of error, due to inter-subject anatomical variability and inter-subject heterogeneity in the location of activations (e.g. Amunts et al., 1999; Fischl et al., 2007; Juch et al., 2005). This variability may moreover affect the conclusions drawn from any group-based approach. Due an imperfect alignment of functional activations between subjects, activations that are located in functionally segregated but adjacent areas at the individual level may be merged into a single functional region at the group level. In the present study, a more reliable conclusion about functional specificity or homology of visually and auditorily-induced vOT activations can be made by employing a subject-specific approach that consists in directly comparing activations observed during reading and speech processing within each individual participant. With this goal in mind, it is now common in cognitive neuroscience to conduct a functional localizer experiment to localize within a participant, the area of interest activated in a given experimental condition and then testing whether this area is also activated in another experimental condition (Brett et al., 2002; Fedorenko et al., 2010; Saxe et al., 2006).

1.1. The present study

The present study addresses the theoretical and methodological issues raised above in an fMRI study. Unlike most existing studies on the topic that used single words as stimuli, here we placed participants in a more natural, yet underexplored, situation where spoken sentences were used. The automaticity issue raised in the current literature was dealt with by orthogonally manipulating a top-down task-driven factor and a bottom-up stimulus-driven factor that have been shown to interfere with processing difficulties. Task demands were manipulated by asking participants to perform low-level perceptual and comprehension tasks. While the former merely required participants to decide whether the same sentence was presented twice in a row, the latter explicitly required extraction of semantic content from spoken inputs. Importantly, neither of them required extraction of orthographic information. The bottom-up stimulus-driven factor corresponded to the quality of spoken inputs that

was manipulated by adding realistic but incomprehensible "multi-speaker" conversation noise to the speech signal on half of the trials. This "cocktail party" situation is common in daily life. Thus, by implementing it in the experimental protocol, we would gain insight into how natural language is processed in the brain.

In order to address the methodological issue regarding the impact of baseline on the emergence of vOT activation, the "multi-speaker" conversation noise was also used as a baseline condition in addition to the classical silent rest baseline. During this "noise-baseline" condition, participants were passively exposed to the "multi-speaker" conversation noise. By contrasting vOT activity obtained in the active (trials with spoken sentences) and each of the two baseline conditions, we would be able to directly examine to what extent the choice of baseline affect our conclusion.

Finally, in addition to spoken sentence processing tasks, participants also underwent a "Visual-vOT" localizer session. This fMRI session allowed us to identify, in individual participants, the voxels within the ventral occipitotemporal pathway that respond to written inputs. These "Visual-vOT" voxels were used as subject-specific functional ROI (fROI). As will be described below, several complementary ROI analyses were performed to explore at the individual level the anatomical overlap between the neural responses induced by the two language modalities.

2. Materials and methods

2.1. Participants

Twenty-four native French speakers participated in the study (mean age: 24.0, range 20–32; 11 women). All were right-handed, with normal hearing and vision and reported no history of neurological or language disorders. They also met all the criteria required to undergo an MRI. They were paid for their participation and gave their written consent, conforming to the Helsinki declaration. The experiment was approved by the local ethics committee (CPP Sud Méditerranée I #RCB 2015-A00845-44).

2.2. Stimuli

The stimuli were 300 spoken sentences presented in French. Two hundred and eighty expressed true statements (e.g., "Le chocolat est produit à partir de cacao", meaning "Chocolate is made from cocoa") while twenty expressed false statements (e.g., "Une longueur se mesure en grammes", meaning "The length is measured in grams"). The latter served as target stimuli in the comprehension task. The sentences were composed of 6.6 words (range: 5–9) and 9.9 syllables (range: 6–14) on average. They were digitally recorded by a native French female speaker using an AKG C1000S microphone in an anechoic chamber, with a sampling rate of 48 kHz (32 bits). The acoustic duration of sentences was 1.7 s on average (standard deviation: 0.27 s). Twenty additional true and one false statements were recorded for practice trials. All sentences were initially selected from a pool of 340 true and 170 false statements that had been judged by an independent group of thirty-five participants. A four-point scale ranging from "absolutely true" (=1) to "absolutely false" (= −1) was used. Sentences selected for the main experiment as true statements received a score of +0.93 on average. Those selected as false statements received a score of −0.97 on average.

Speech-in-noise stimuli were constructed by embedding the spoken sentences in a "multispeaker babble" (MSB) noise, at a fixed signal-to-noise ratio (SNR) of +6 dB. The SNR value was chosen based on the results of a preliminary experiment conducted on an independent group of twelve participants, out of the scanner but in the presence of recorded gradient noise, aiming at determining the effects of the amount of noise on the intelligibility of the speech signal. The mean intelligibility score of the speech-in-noise sentences at +6 dB was 91.9% on average (see supplementary methods). MSB was constructed by superimposing auditory pseudo-sentences (sentences composed of pseudo-words) recorded by six native French speakers (3 women, 3 men) at the sampling rate of 48 kHz, 32bits. For each speaker, two different tracks, each composed of a succession of 25 pseudo-sentences presented in a random order, were constructed. The resulting 12 tracks were then superimposed to obtain a 1 min-long wav file of realistic, dense, MSB. To construct the speech-in-noise version of each sentence, a segment of MSB of the same duration as the sentence was randomly selected from the 1 min-long MSB wav file. The intensity of the MSB segment was rescaled before superimposing it on the sentence to achieve the desired SNR of +6 dB.

Twenty additional wave files containing only MSB noise were constructed and served in a control condition, in which the acoustic properties of real speech are preserved while eliminating any semantic contents. Their average duration matched the average duration of the spoken sentences (1.705 s on average, standard deviation: 0.33 s). The root-mean-square (RMS) amplitude was equalized across stimulus types (spoken sentences, spoken sentences embedded in MSB, MSB only). All signal processing was performed in Matlab R2015a (Mathworks Inc., Natick, MA, USA).

2.3. Spoken sentence processing: tasks and procedure

Participants performed a perceptual task and a sentence comprehension task in the scanner. In the perceptual task, they had to press the response button when a sentence was repeated twice in a row (Go trials). In the comprehension task, they had to detect false statements (Go trials). While the first task relied on low-level perceptual process, efforts on comprehension were crucial in the second task. Each task contained 20 Go trials and 140 No-Go trials, the latter always corresponding to true statements. Half of the sentences were presented against clear background (*clear speech condition*) and half were presented against MSB noise at an SNR of +6 dB (*speech-in-noise condition*). Each participant heard each No-Go sentence only once. Across participants, each No-Go sentence appeared equally in the four listening conditions defined by the combination of task (perception vs. comprehension) and quality of speech signal (clear vs. noise). After the practice trials, the two tasks were presented alternately in four runs of 7.2 min each (two runs per task). Each run contained 10 Go trials that were distributed in a pseudo-random manner amongst 70 No-Go trials (there was at least one No-Go trial between two Go trials). These 80 "active" trials were grouped into 20 short blocks of four trials. In addition, 10 rest blocks were included, half corresponded to silent background and the other half corresponded to MSB noise without sentence. Active and rest block duration was 14s on average (range 12s–18s). Run order was counterbalanced across participants, and blocks of different conditions were presented in a pseudo-randomized order within each run. Two consecutive blocks of the same condition were avoided (see Fig. 1A). Within each trial, a sentence (duration varying from 1s to 2.4s) was presented with a visual fixation cross remaining on the screen for the same duration. This was followed by a blank screen whose duration was jittered so that the SOA (3.55s on average) followed an exponential curve to maximize design efficiency (see Henson, 2015). The same trial sequence was used during the rest trials where sentences were replaced by silence or MSB noise.

2.4. "Visual-vOT" localizer

The main objective of the present study was to examine whether the vOT, described in the literature as primarily involved in visual word processing, is activated during spoken sentence processing. In order to localize the voxels within the area that respond to visual words ("visual-vOT"), at the individual level, participants underwent an additional functional localizer run that lasted 7.4 min. Single words and consonant strings were visually presented in short blocks. Each block contained a sequence 24 stimuli of the same category. Each stimulus remained on the screen during 340 ms, followed by a blank screen of variable duration (mean 160 ms). This led to a block duration of 12s on average (range 11s–13.3s). The run included 24 "active" blocks (12 per stimulus

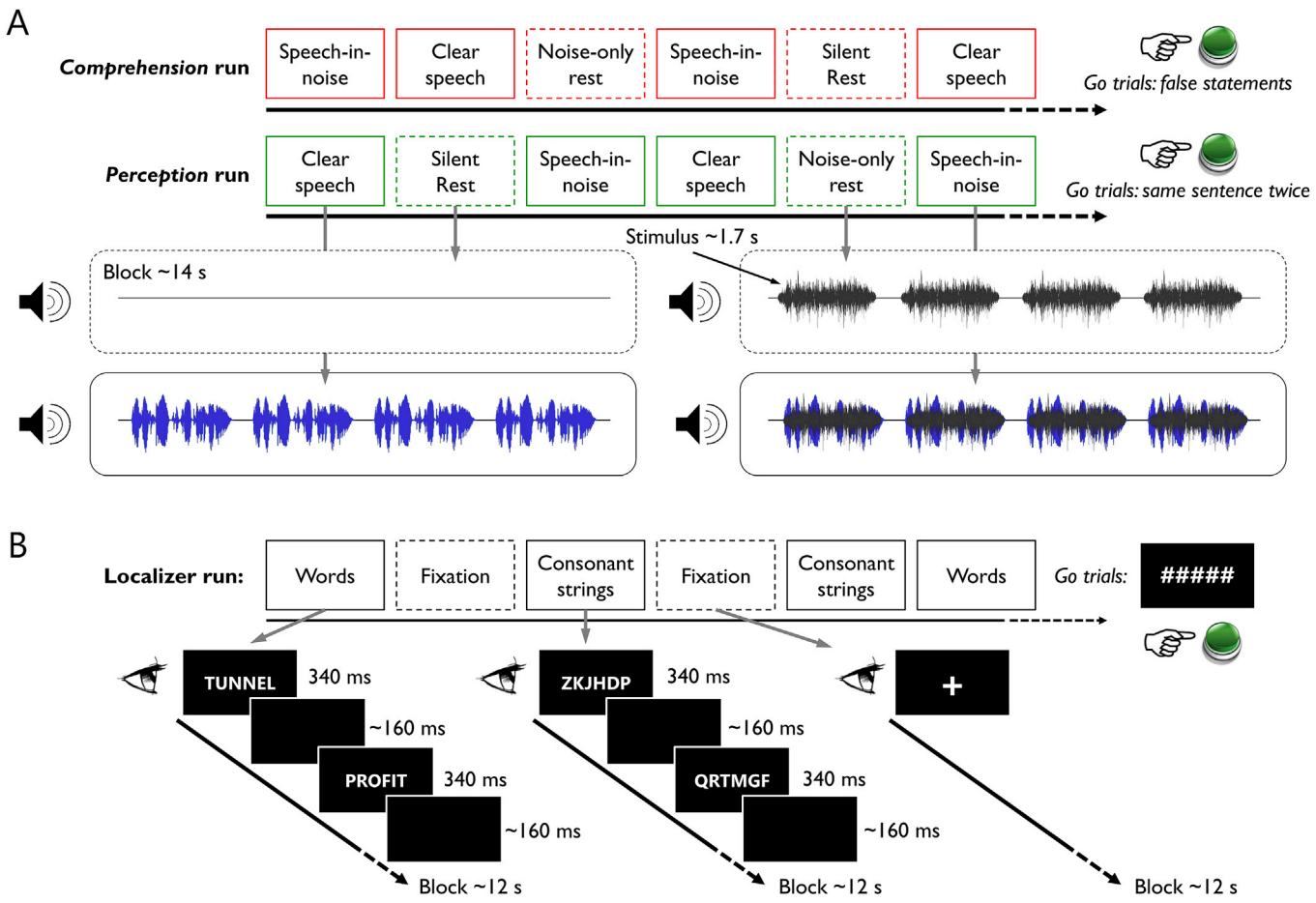


Fig. 1. (A) Task design for the speech processing experiment. Each block (except "silent rest") contained four stimuli of the same type, i.e., sentences (clear speech blocks), sentences embedded in conversation noise (speech-in noise blocks) or conversation noise ("noise-only rest" blocks). The comprehension and the perceptual tasks were presented in separated runs (two runs per task); each containing the four types of blocks presented in a pseudo-random order (30 blocks/run). Within each task, 20 Go trials were randomly presented in the "active" blocks. They corresponded to false statements in the comprehension task and to the same sentence presented twice in a row in the perception task. (B) Task design for the "visual-vOT" localizer experiment. Three types of stimulus, i.e., words, consonant strings and fixation cross were presented in a block design. Each "active" block contained 24 stimuli of the same type, i.e., either words or consonant strings. Twelve blocks of each stimulus type were presented in a pseudo-random order. Twelve Go trials corresponding to hash symbols were randomly presented in the "active" blocks.

category) and 12 "passive" blocks where a fixation cross remained on the screen throughout the block duration (see Fig. 1B). Twelve target stimuli ("######") randomly appeared on the screen throughout the run. Participants had to detect them by pressing the response button. All stimuli were presented in the center of the screen, in white font on a dark-grey background. Word stimuli corresponded to 282, 6-letters long, monosyllabic and disyllabic, nouns and adjectives, of moderate lexical frequency (7.21 per million, on average). They were selected from the French database LEXIQUE (New et al., 2004). Consonant strings corresponded to 282 randomly generated, unpronounceable, sequences of six consonants.

2.5. MRI data acquisition

Data were collected on a 3-T Siemens Prisma Scanner (Siemens, Erlangen, Germany) at the Marseille MRI center (Centre IRM-INT@CERIMED, UMR7289 CNRS & AMU) using a 64-channel head coil. 353 functional volumes covering the whole brain were acquired during each of the four runs of the speech processing experiment and 363 functional volumes were acquired during the localizer experiment, both using the same BOLD-sensitive gradient EPI sequence (TR = 1224 ms, echo time = 30 ms, flip angle = 66°, 54 slices with a thickness of 2.5 mm, FOV = 210 × 210 mm^2, matrix = 84 x 84, slice thickness = 2.5 mm, multiband factor = 3). Prior to functional imaging, T1-weighted (MPRAGE sequence, voxel size = 1 x 1 × 1 mm3, data matrix 256 x 256, TR/TI/TE = 2300/900/2.98 ms, flip angle = 9°) and FieldMap (Dual echo Gradient-echo acquisition with TR = 677 ms, TE1/TE2 = 4.92/7.38 ms and FOV = 210 × 210 mm2, voxel size: 2.2 x 2.2 × 2.5 mm3) images were acquired for offline preprocessing procedure (e.g., unwrapping, normalization). Auditory hardware channel was composed by the Sensimetrics S14 MR-compatible insert earphones completed with a Yamaha P-2075 power amplifier. Both auditory and visual stimuli were managed and delivered using an in-house software using the NI LabVIEW environment. The software was launched and real-time synchronized with the MR acquisition using a NI-PXI 6289 digital input/output hardware, which also allowed behavioral responses recording.

2.6. fMRI data pre-processing and analysis

Preprocessing and statistical analyses were performed using SPM12 (Wellcome Trust Centre for Neuroimaging, University College London, London, UK) running in Matlab R2015a (Mathworks Inc., Natick, MA, USA). All functional images were slice-time corrected, unwarped using the FieldMap toolbox, realigned to the mean of the images, normalized in Montreal Neurological Institute (MNI) space using the deformation field generated during segmentation of the high-resolution structural image, and smoothed with an 5-mm FWHM Gaussian kernel (while retaining the

original voxel size). A detection of global mean intensity and motion outliers was performed using Artifact Detection Tools (ART, http://www.nitrc.org/projects/artifact_detect/). Marked outliers represented 0.3% of all acquired volumes (displacement larger than 2 mm from one volume to the next, using the norms of the linear motion parameters and of the angular motion parameters). Three participants were excluded from all analyses, two because of technical issues leading to poor listening conditions, and one because of excessive head movements during the acquisition (translation of more than 6 mm in one direction).

In the spoken sentence processing tasks, the first-level General Linear Model (GLM) comprised four task regressors (i.e., comprehension task with clear speech, comprehension task with speech-in-noise, perceptual task with clear speech, perceptual task with speech-in-noise), in which stimuli were modeled as events of variable duration to account for the differences in acoustic duration of the stimuli and four regressors for the rest conditions (i.e., silent rest and noise-only rest during comprehension task, silent rest and noise-only rest during perceptual task). An additional regressor was included to account for behavioral responses (as events with duration of 0). Regressors were convolved with the canonical hemodynamic response function (HRF), and the default SPM autoregressive model AR(1) was applied. Functional data were filtered with a 128s high-pass filter. Six motion regressors, one regressor for session and, when applicable, outlier regressors from the ART procedure were included as covariates of non-interest. Statistical parametric maps for each of the eight experimental conditions (i.e., four for tasks and four for rest conditions) and each participant (beta maps) were calculated at the first level, and were entered in a second-level within-subjects one-way analysis of variance (random effects analysis). Unless stated otherwise, all statistical comparisons were performed with a voxelwise threshold of $p < .001$ and a cluster extent threshold of $p < .05$ FWE-corrected. A different GLM was computed for the visual-vOT localizer task, with the word, consonant string and fixation conditions modeled as blocks. One regressor for behavioral responses and six motion regressors were also included in the design matrix. Similar procedures and thresholds as in the spoken sentence processing experiment were applied for statistical analyses.

2.7. ROI analyses

In order to specifically examine the involvement of the left vOT in speech processing, two complementary approaches were used. In the first, namely the *literature-based ROI approach*, the ROIs corresponded to six sub-regions along the posterior to anterior portions of the vOT. The volumes-of-interested were specified based on the coordinates from the literature on visual word processing. Precisely, six 6-mm radius spheres, positioned along the ventral occipitotemporal pathway, were built based on the coordinates reported by Vinckier et al. (2007): ROI 1: 18 -96 -10, ROI 2: 36 -80 -12, ROI 3: 46 -64 -14, ROI 4: 48 -56 -16, ROI 5: 50 -48 -16 and ROI 6: 50 -40 -18 (Fig. 8A). For each ROI, subject-specific contrast estimates were extracted for the contrasts of interest from both speech processing (perceptual task with clear speech > silent-rest, perceptual task with speech-in-noise > silent-rest, comprehension task with clear speech > silent-rest, comprehension task with speech--in-noise > silent-rest) and visual word processing experiments (visual words > consonant strings). One-sample t-tests were used to test for significant activations. In the speech processing experiment, p values obtained at each ROI were adjusted for multiple comparisons, using Bonferroni correction.

The literature-based ROI approach was complemented by the subject-specific ROI approach in which ROIs were defined based on subjects' functional data corresponding to the vOT voxels activated by reading written words (i.e., fROI). More specifically, the fROIs were built by intersecting each individual, first-level, functional map (from the visual-vOT localizer contrast) with a search volume corresponding to the vOT cluster of the second-level group functional map ($p < .001$, uncorrected). Three complementary sets of analyses were conducted in order to examine 1) whether the vOT voxels activated during speech processing (henceforth, "auditory-vOT") were those activated during visual word processing ("visual-vOT"), 2) the degree of overlap between these visual-vOT and auditory-vOT voxels and 3) the correlations between the activation patterns obtained in the speech processing and visual word processing experiments. To facilitate the comprehension of the results, more extensive details on the analyses will be presented in the Results section.

3. Results

3.1. Behavioral results

Each task included 20 target stimuli (Go trials). Half of them were presented against clear background and the others against MBS background. Targets were false statements in the comprehension task and sentences repeated twice in a row in the perceptual task. The average hit rate was high: 85%. A repeated-measures ANOVA, with task and presence of noise as within-subject factors, showed that hit rates varied significantly across experimental conditions (Fig. 2). Hit rate was higher during the perceptual task compared to the comprehension task (94% vs. 76%; $F(1, 20) = 36.0$, $p < .0001$; −18, with 95% confidence interval CI = −25, −12), and higher with clear speech compared to speech-in-noise (92% vs. 78%; $F(1, 20) = 25.7$, $p < .0001$; −15 with 95% CI = −20, −9). The interaction was also significant ($F(1, 20) = 13.9$, $p < .002$). Post-hoc Scheffé's test indicated that the effect of noise was significant in the comprehension task (88% for clear speech vs. 63% for speech-in-noise; $p < .0001$; −25 with 95% CI = −37, −13) but not in the perceptual task (96% vs. 92%; $p = .75$; −4 with 95% CI = −16, 8). Average false alarm rate was low (1.2%) and was modulated only by task: 0.4% for the perceptual task as opposed to 2.0% comprehension task ($F(1, 20) = 35.2$, $p < .0001$; +1.6 with 95% CI = 1.0, 2.2). A repeated-measures ANOVA was also performed on reaction times for correct responses (Fig. 2). It revealed both main effects of task (1354 ms for perception vs. 2482 ms for comprehension; $F(1, 20) = 122.1$, $p < .0001$; +1128 ms with 95% CI = 915, 1340) and noise (1842 ms for clear speech vs. 1994 ms for speech-in-noise; $F(1, 20) = 20.3$, $p < .0003$; +152 ms with 95% CI = 82, 223) without a significant interaction between the two factors ($F(1, 20) = 1.13$, $p = .30$).

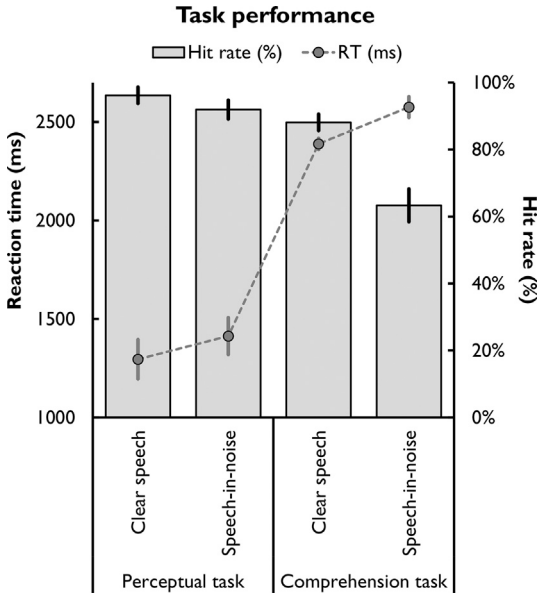


Fig. 2. Behavioral results. Average reaction time for correct trials (left y-axis, dots and dashed line) and average hit rate (right y-axis, bar graph) obtained in the perceptual task and the comprehension task when the stimuli were presented against clear and MBS noise background. Error bars represent SEM.

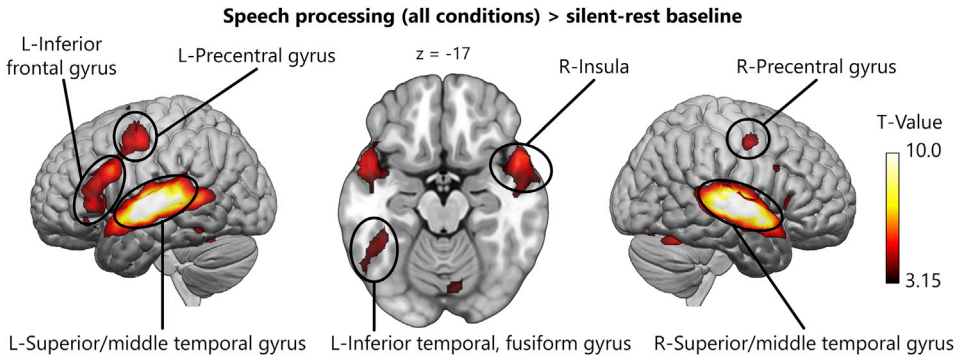


Fig. 3. Activations for all speech processing conditions combined against silent rest baseline (voxelwise $p < .001$ unc. and at the cluster-level $p < .05$ FWE-corr).

3.1.1. Whole brain analyses

Whole-brain analyses were conducted in order to characterize the global network of brain areas recruited during spoken sentence processing as well as the effects of task demands, the presence of noise in speech signal and the interaction between these two factors. Then, we further examined brain responses obtained in each of the four speech processing conditions, focusing on the presence of vOT activations.

3.1.1.1. Global speech processing network. In order to characterize the global network of brain areas recruited during spoken sentence processing, activations observed in the four speech processing conditions were combined and contrasted against silent rest baseline. Extensive activations were found, predominantly in a bilateral temporal region (superior and middle temporal gyri, Heschl gyrus), the left inferior frontal gyrus (pars triangularis and opercularis) and the left precentral gyrus (see Fig. 3), thus suggesting an involvement of the phonological, semantic and articulatory systems. Smaller significant clusters were found in the bilateral insula, the supplementary motor area (SMA), the cerebellum, the right precentral gyrus, the right inferior frontal gyrus, the left hippocampus, the bilateral cuneus. Importantly, a significant activation was observed in the left vOT (inferior temporal gyrus, fusiform gyrus, peak at −43, −44, −15).

3.1.1.2. Task effect. When the perceptual task was contrasted against the comprehension task ([perceptual task with clear speech + perceptual task with speech-in-noise] – [comprehension task with clear speech + comprehension task with speech-in-noise]), two significant clusters were found in the anterior part of the left middle frontal gyrus (MFG) and in the right precuneus (see Supplementary Table S1). The opposite, comprehension task > perceptual task, contrast yielded eight significant clusters (see Fig. 4 and Supplementary Table S1) located in the bilateral occipital cortex (cuneus, calcarine gyrus, lingual gyrus), subcortical regions (bilateral striatum, thalamus, brainstem) extending into left hippocampus and parahippocampal gyrus, the bilateral inferior frontal gyrus, the supplementary motor area and the right cerebellum. More related to our purpose, a significant activation was also found in the left vOT (temporal inferior and fusiform gyrus; composed of different peaks: at -35 -59 -7; -38 -51 -10 and -48 -56 -15) which indicates that its activation, already observed in the global network of spoken sentence processing, was stronger in high-level comprehension compared to low-level perceptual task. Most of the observed activations are consistent with concept retrieval or selection of semantic information (pars orbitalis, triangularis, hippocampus and parahippocampal gyrus; Badre and Wagner, 2007; Binder et al., 2009). There is no evidence that the level of activation of the core phonological and semantic areas within the global speech network, such as the left superior and middle temporal gyri, depends on task demands. Although a strong conclusion could not be drawn from a null effect and further research is still needed, the absence of task effect suggests a possibility that, up to a certain point, the brain may automatically processes phonological and semantic information contained in meaningful spoken inputs regardless of task demands. Finally, the finding of bilateral posterior medial occipital activation is unexpected. Since these regions are typically involved in visuo-spatial imagery and their activations have been reported in memory retrieval tasks (Burianova and Grady, 2007; Whittingstall et al., 2014) it is possible that for some sentences (e.g., "The bumblebee is bigger than the mosquito", "A square is a four-sided figure") making semantic judgment might have induced visual imagery. It should also be noted that our stimuli that contained a large proportion of concrete, highly imageable words might also have induced activity in the visual cortex to some extent (Fiebach and Friederici, 2004).

3.1.1.3. Noise effect. The speech-in-noise > clear speech contrast ([perceptual task with speech-in-noise + comprehension task with speech-in-noise] – [perceptual task with clear speech + comprehension task with clear speech]) revealed eleven significant clusters (see Fig. 5A and Supplementary Table S2). The largest cluster (1329 voxels) was located in the right insula (peak: 33, 22, 6) and extended to the right pars triangularis (43, 22, 8) and the right middle frontal gyrus (MFG) (40, 47, 31). The second largest cluster of activation (−60, −29, 8) was found in a left superior temporal gyrus (STG) cluster (355 voxels; but "sparing" the left Heschl's gyrus). Other significant clusters were found in the right STG, the left insula (extending into pars triangularis), SMA, the left cerebellum, the right supramarginal gyrus, the right precuneus and the right middle cingulate gyrus. A small cluster was also present in the left inferior frontal gyrus (IFG) (−45, 12, 23), including voxels in pars triangularis and opercularis. Activity of most areas within this network (i.e., STG, SMA, IFG, Insula) have been repeatedly reported in experiments manipulating speech intelligibility using various types of speech degradation methods (Adank and Devlin, 2010; Davis et al., 2011; Obleser et al., 2007; Wild et al., 2012). Interestingly, considering the two tasks together, adding noise in the speech signal did not lead to an increase of activity in the left vOT.

Fourteen clusters showed significant activation in the opposite, clear speech > speech-in-noise contrast (see Fig. 5B and Supplementary Table S2). Most of them were located in areas involved in storage and retrieval of semantic knowledge (Binder et al., 2009). The largest cluster was located in the left middle occipital/angular gyrus (AG) (−40, −71, 33 and −48, −61, 26) and extended into to the posterior part of the left middle temporal gyrus (MTG) (−55, −44, −7). AG activation is consistent with facilitated sentence comprehension when speech is clearly presented. The literature indicates that this area is particularly involved in processing concepts (Seghier, 2013) and its activation seems to be affected by the quality of speech signal (Obleser and Kotz, 2010). Bilateral activation was observed in several brain areas including: a parietal superior and inferior region (postcentral and precentral gyri), the rolandic operculum (extending into the Heschl's gyrus in the right hemisphere), the insula, the fusiform gyrus (medial), the amygdala, the

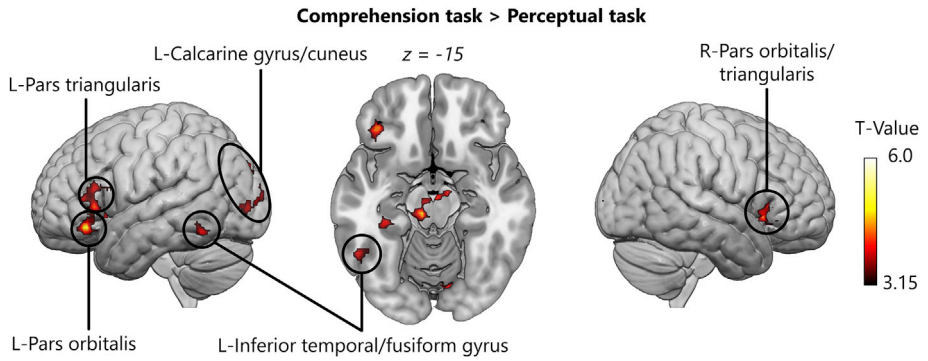


Fig. 4. Brain areas showing higher activation during the comprehension compared to the perceptual task (voxelwise $p < .001$ unc. and cluster $p < .05$ FWE-corr).

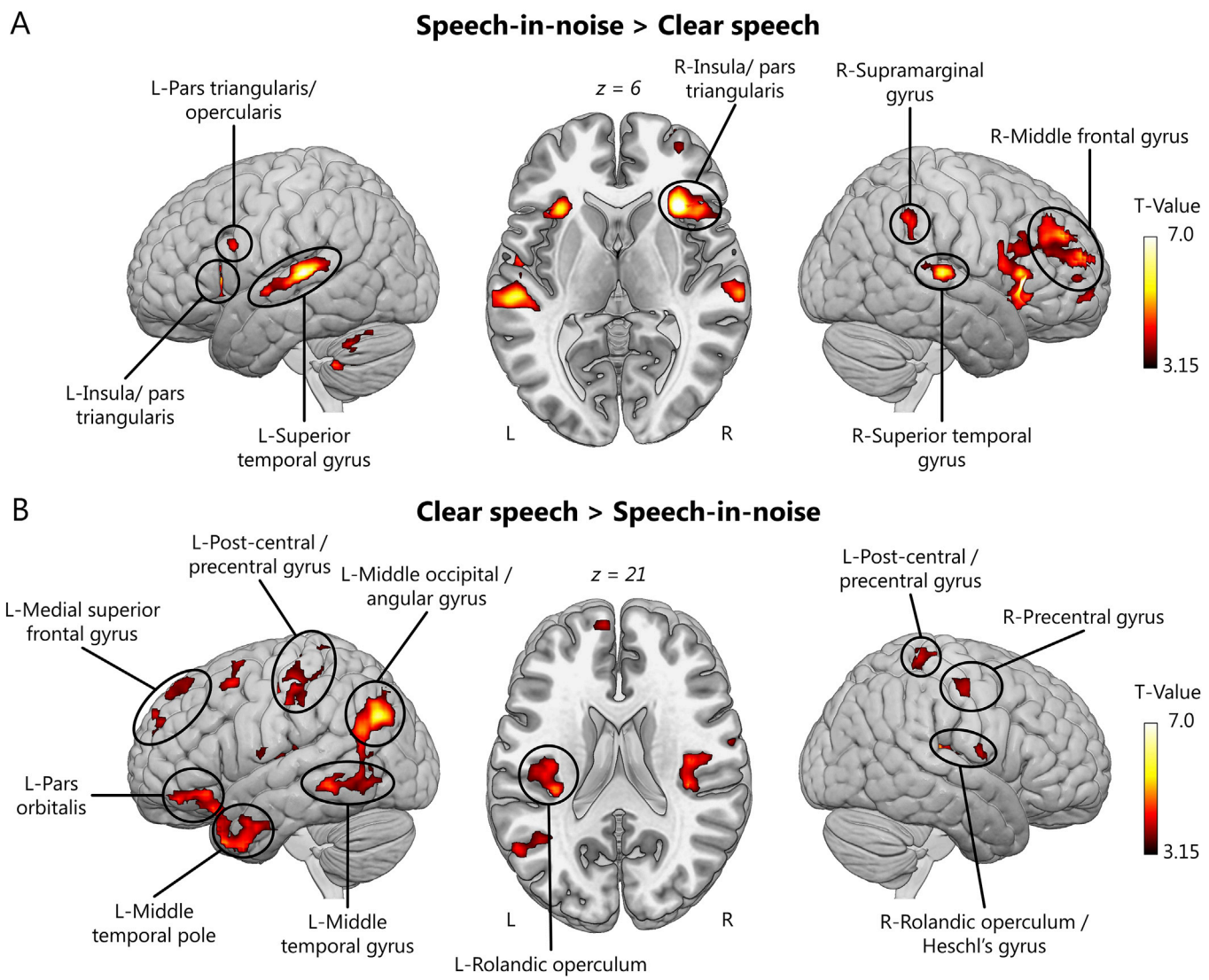


Fig. 5. (A) Activations obtained in the speech-in-noise > clear speech contrast (B) Activations obtained in the clear speech > speech-in-noise contrast (voxelwise $p < .001$ unc. and cluster $p < .05$ FWE-corr).

hippocampus, the putamen, the posterior and middle cingulate gyrus, and the precuneus. The activation observed in the pars orbitalis, MFG and the medial superior frontal gyrus and the middle temporal pole was left lateralized. The increase of activity within the temporal pole, considered as a "semantic hub", further supports the idea that semantic processing is facilitated in the clear speech condition. No activation was found in the ventral occipitotemporal cortex.

3.1.1.4. Interaction between task and noise. The interaction between the effect of task and the presence of noise was examined via two opposite T-contrasts: The first contrast aimed at identifying the areas where the presence of the background noise reduced brain responses during the

comprehension task but enhanced responses during the perceptual task (contrast weights [+1 -1 -1 +1]: the first two digits represent the two comprehension conditions where clear speech and speech-in-noise were presented, respectively. The last two digits represent the two perception conditions where clear speech and speech-in-noise were presented, respectively). The second contrast aimed to identify the areas where the presence of the background noise enhanced responses during the comprehension task but reduced responses during the perceptual task (contrast weights [-1 +1 + 1 -1]). While the latter contrast did not reveal any significant activation, the former one led to significant brain activity in ten clusters (see Fig. 6 and Supplementary Table S3). Most activated voxels were included in a large motor/premotor bilateral cluster (6077 voxels), covering bilateral postcentral and precentral gyri from the ventral part up to the SMA, and extended into subcortical regions (putamen, caudate nucleus, thalamus amygdala) and insula. This involvement of the sensorimotor cortex during speech perception can be explained in the context of the motor theory of speech perception, according to which information on articulatory features is accessed during speech processing (Liberman and Whalen, 2000; Pulvermüller et al., 2006). The interaction depicted here further suggests that the contribution of the articulatory system to speech processing may not be systematic. As shown in the present finding, its contribution was reduced when speech signal was degraded, especially in a demanding task. The same activation pattern was also found in the right fusiform/inferior temporal gyrus, left pars triangularis (−33, 37, 3), right orbito-frontal cortex, medial part of the superior frontal gyrus, right MFG, cerebellum and, interestingly, the left vOT (fusiform/temporal inferior gyrus; peak at −45, −49, −25).

3.1.1.5. Left vOT involvement in different speech processing contexts. The involvement of the left vOT in the four speech processing contexts was examined by contrasting the brain activity obtained in each of the four experimental conditions against the silent rest baseline. Using the same statistical threshold as in the above analyses, we observed a similar pattern of activity as in the global speech processing network, i.e., significant activations were found in most brain areas that are part of the phonological, articulatory and semantic networks. Crucially, activations within the vOT were found in all conditions except in the perceptual task when sentences were presented against background noise. The extent and strength of activity within the vOT cluster varied greatly across conditions. As shown in Fig. 7A, the activation is strongest and most widespread in the comprehension task with clear speech. It became smaller in the perceptual task. Based on existing literature, the coordinates of vOT observed in our speech processing tasks (peak at x = −43, y = −44, z = −15 when all tasks were combined; cf. Fig. 3) are in the vicinity of the area reported to be involved in visual word processing (for instance, coordinates x = −44, y = −58, z = −15 in a meta-analysis by Jobard et al., 2003).

As stated earlier, previous studies showed that the detection of vOT activation in some speech processing tasks depended on the choice of baseline, with "non-language auditory" baselines being more effective in uncovering vOT activation than silent rest or implicit baseline (e.g., Ludersdorfer et al., 2016). This pattern was replicated in the current study. Indeed, when noise-only rest was considered as baseline, the level of vOT activation observed during spoken sentence processing increased and became significant in all conditions (Fig. 7B). As shown in Fig. 7C, this was explained by a deactivation of the left vOT observed during the noise-only rest (compared to implicit baseline), but not during silent rest. Since the use of noise-only rest baseline artificially boosted the significance level of vOT activation, in order to be conservative, only the silent rest baseline was applied in the following analyses.

3.1.2. ROI analyses

The two ROI analyses described below relied on the vOT activation obtained in the localizer experiment contrasting written words to consonants strings. These results were used to explore the activation pattern along the ventral occipitotemporal pathway (see *Approach 1 below*) and to examine whether the voxels activated by written words were also involved in speech processing (see *Approach 2 below*).

3.1.2.1. Preliminary step: visual-vOT localizer experiment results. Behavioral data showed that all of the target stimuli (######) were correctly detected by all participants. There were no false alarm and reaction times were similar for targets that appeared in the word and the consonant string blocks (484 ms vs. 467 ms, respectively; $F(1,20) = 3.04$; $p = .097$; −16 ms, with 95% CI = −3, 36). Regarding neuroimaging results, three clusters of similar extent (around 265 voxels) showed significant activation in the words > consonant strings contrast (using a voxelwise threshold of $p < .001$ with a cluster threshold of $p < .05$ FWE-corrected at the whole brain level; see Supplementary Table S4). The first cluster was located in the left middle temporal gyrus (MTG) (−60, −34, 3), the second in the left vOT (highest peak at −43, −44, −17, see Fig. 7D) and the third in an inferior frontal region stretching from the postcentral gyrus (−55, −6, 46) to the pars opercularis of the IFG (−48, 9, 16). The peak coordinates of the vOT cluster found in this experiment corresponds to the one found in the speech processing experiment (−43,

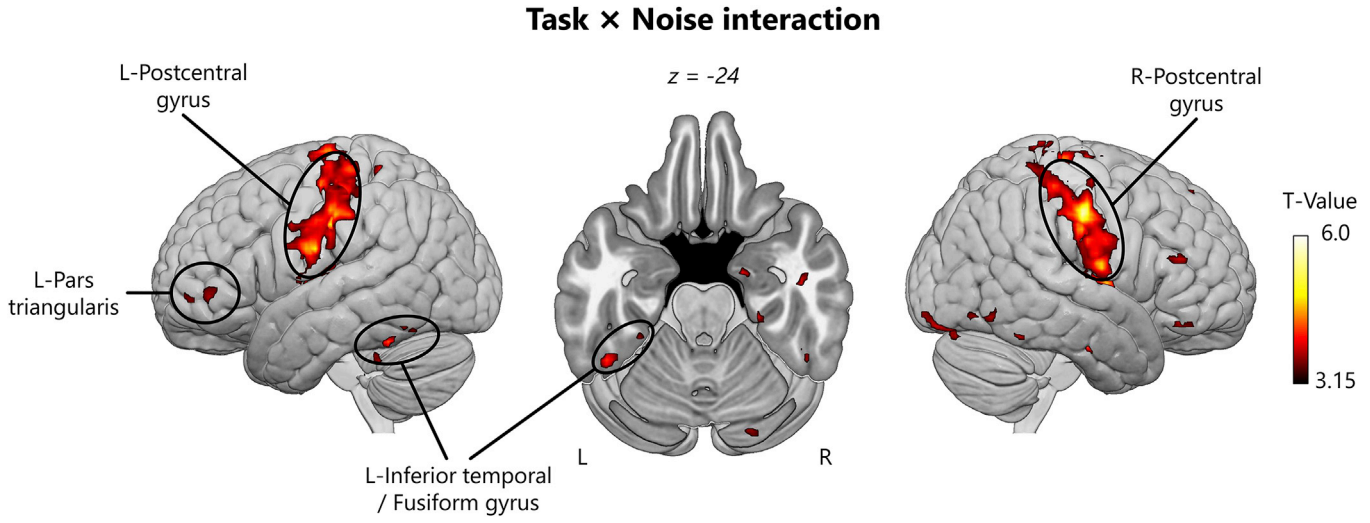


Fig. 6. Brain areas showing a stronger reduction of activity due to the presence of noise in the comprehension task compared to the perceptual task (contrast weights [+1 -1 -1 +1], see text) (voxelwise $p < .001$ unc. and cluster $p < .05$ FWE-corr).

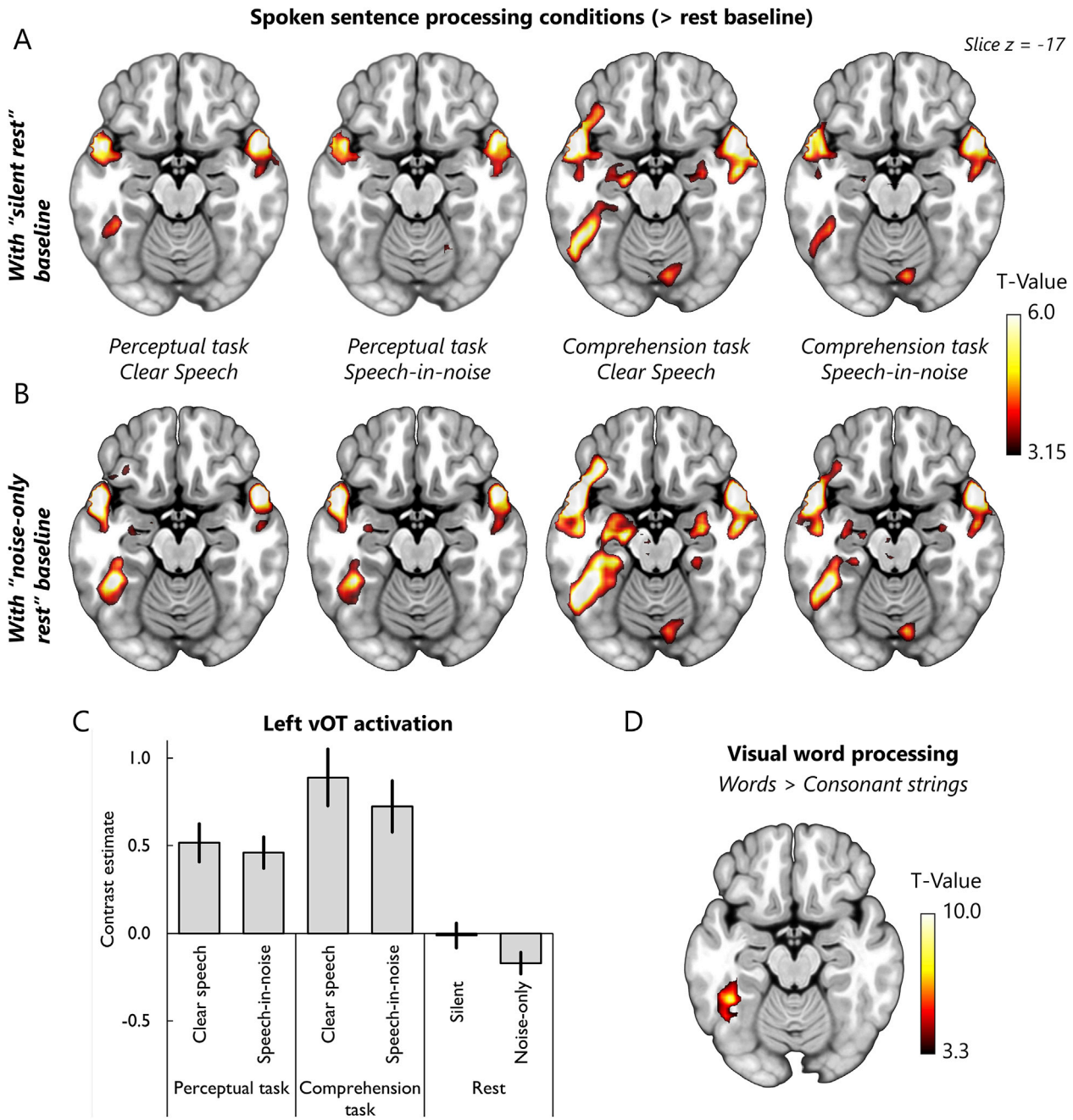


Fig. 7. (A) Brain responses obtained in the four experimental conditions from the spoken sentence processing experiment compared to "silent-rest" baseline and (B) compared to "noise-only rest" baseline. (C) Average contrast estimates for the four speech processing conditions and the two rest conditions (against implicit baseline), at the left vOT activation peak (i.e., −43, −44, −15). Error bars represent SEM. (D) Words > consonant strings contrast from the visual-vOT localizer experiment, showing a left vOT activation cluster (voxelwise $p < .001$ unc. and cluster $p < .05$ FWE-corr for all activation maps).

−44, −15). This is a first indication that the same area seems to be recruited in both language modalities.

3.1.2.2. Approach 1: literature-based vOT coordinates. One important property of the visuo-orthographic pathway involved in reading is its hierarchical organization. An increasing sensibility to word-like visual stimuli along a posterior-to-anterior direction in the left vOT is documented in the literature (Dehaene et al., 2005; Vinckier et al., 2007). Similar to Ludersdorfer et al. (2013), we examined if a posterior-to-anterior gradient of activation in the ventral visual pathway could also be observed in the speech processing conditions presented here. As shown in Fig. 8, in the six posterior-to-anterior ROIs described by Vinckier et al. (2007), we observed a similar activation profile for spoken sentence and visual word processing. Overall, the activity increases progressively from the posterior ROIs to reach the highest activation in ROIs 4 ($z = -56$) and 5 ($z = -48$), then the activation decreases in the most anterior part of the pathway. One-sample T-tests (with Bonferroni corrections applied for multiple comparisons) performed on the results from the speech processing experiment (Fig. 8B) indicated that, in the perceptual task, the activity was significantly higher than in the silent rest condition at ROI 5 both in the speech-in-noise ($t(20) = 3.24$, $p.corr < 0.02$) and in the clear speech conditions ($t(20) = 4.30$, $p.corr < 0.002$). The activation obtained in the latter condition almost reached significance at ROI 4 ($t(20) = 2.70$, $p.corr = 0.082$). In the comprehension task, significant results were observed at ROIs 4 and 5 both in the speech-in-noise ($t(20) = 5.44$,

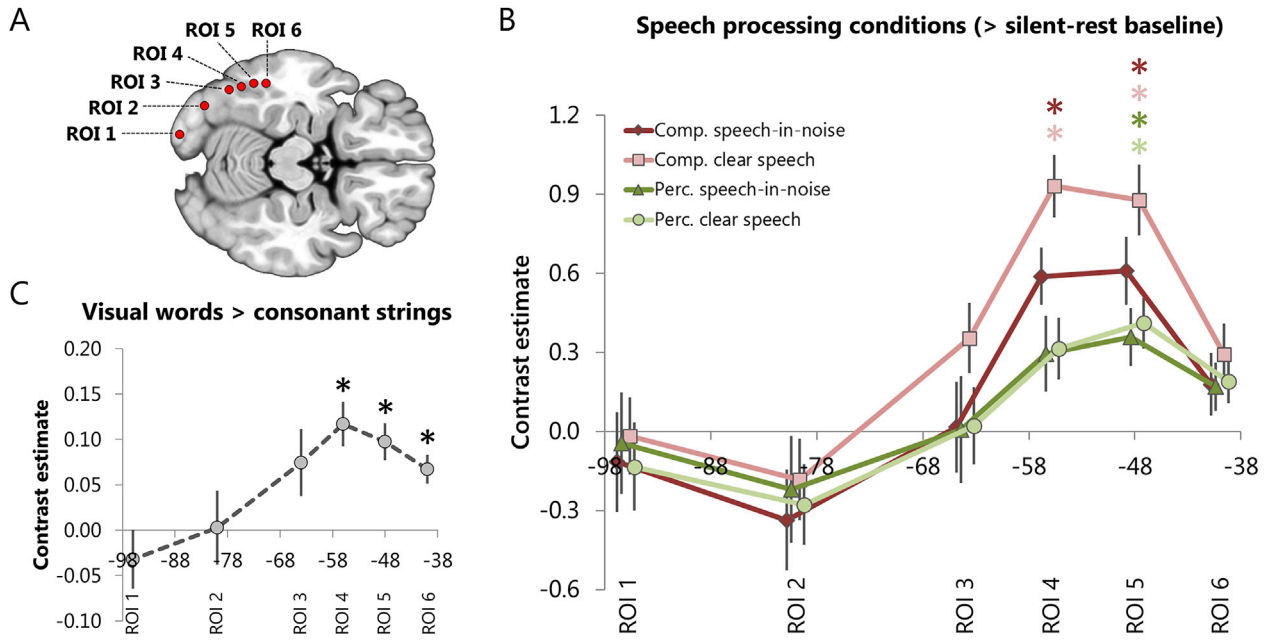


Fig. 8. (A) Locations of six ROIs from Vinckier et al. (2007) displayed in the axial plane: ROI 1 = −18 -96 -10, ROI 2 = -36 -80 -12, ROI 3 = −46 -64 -14, ROI 4 = -48 -56 -16, ROI 5 = -50 -48 -16 and ROI 6 = -50 -40 -18. (B) Posterior-to-anterior gradient of vOT activation obtained in the six 6-mm radius spherical ROIs in each of the four "task > silent rest" contrasts of the spoken sentence processing experiment. (C) Posterior-to-anterior gradient of vOT activation obtained in the six 6-mm radius spherical ROIs in the "visual words > consonant strings" contrast of the visual-vOT localizer experiment. X-axis represents the z-coordinate (slightly jittered for visualization purposes). * indicates the contrasts showing significant results.

p.corr < 0.0003 and t(20) = 4.75, p.corr < 0.0008, for ROIs 4 and 5, respectively) and in the clear speech conditions (t(20) = 7.86, p.corr < 0.0001; t(20) = 6.55, p.corr < 0.0001, for ROIs 4 and 5, respectively). Interestingly, activation extracted from theses ROIs varied with experimental conditions: A 2 × 2 repeated-measures ANOVA conducted on the contrast estimates extracted at these ROIs showed a significant interaction between task and presence of noise ($F(1, 20) = 10.71$; $p < .004$; $F(1, 20) = 6.61$; $p < .018$, for ROIs 4 and 5 respectively). In line with the findings obtained in the whole brain analyses, the presence of noise reduced the activation in the comprehension (post-hoc Sheffe's test, both ps < .02) but not in the perceptual task (both ps > .80).

The analysis performed on the results from the localizer task (Fig. 8C) showed that the visual words > consonant strings contrast led to significant activation at ROIs 4, 5, and 6 (t(20) = 4.79, p.corr < 0.0002; t(20) = 4.78, p.corr < 0.0002; and t(20) = 4.24, p.corr < 0.0004, respectively). This activation profile is fully consistent with the literature-based location of the VWFA (i.e., around y = −57) (Cohen et al., 2000, 2002). Taken together, the activations induced by the two language modalities follow the same pattern along the ventral occipito-temporal pathway. Additionally, the peak activations observed during speech and written word processing were located in the same portions of the pathway (ROI 4, ROI 5).

3.1.2.3. Approach 2: subject-specific functional ROIs. Based on the group results reported above, one could reasonably assume that the area within the ventral occipitotemporal pathway that was active during sentence processing ("auditory-vOT") was also active when participants processed written words ("visual-vOT"). However, as mentioned in the Introduction section, due to inter-individual anatomical variability, an apparently identical and homogeneous activation location in the auditory and written language tasks at the group level may result from the possibility that distinct brain responses at the individual level were averaged across participants. Moreover, meaningful activation may have been missed if functionally equivalent regions were misaligned across individuals. In the following analyses, we adopted a subject-specific approach using functionally defined ROIs (fROIs). Three complementary sets of analyses were conducted to address three questions:

1) Were the vOT voxels activated during visual word processing also activated during speech processing?

In the first set of analyses, the subject-specific ROI was defined as the intersection between "group search volume" (i.e., the vOT cluster observed in the group analysis of the visual words > consonant strings contrast; 275 voxels, Fig. 7D) and individually identified voxels activated in the same contrast, using voxelwise, p < .001, uncorrected threshold. Four participants who did not show any significant voxel within the group search volume were excluded from further analyses. The average number of voxels in this subject-specific functional ROI was 49 (SD = 38). For each participant, we extracted the contrast estimates for the four speech processing conditions (contrasted against the silent rest baseline) within his/her functionally defined ROI. As illustrated in Fig. 9, one sample T-tests (with Bonferroni corrections for multiple comparisons) showed significant activations in the four speech processing conditions: perceptual task performed on clear speech (t(16) = 10.13, p.corr < 0.0001) and on speech-in-noise (t(16) = 9.31, p.corr < 0.0001), comprehension task performed on clear speech (t(16) = 9.02, p.corr < 0.0001) and on speech-in-noise (t(16) = 8.61, p.corr < 0.0001). A 2x2 repeated-measures ANOVA showed significant task effect (1.38 for comprehension task vs. 0.90 for the perceptual task; $F(1, 16) = 16.17$; $p < .001$), noise effect (1.24 for clear speech vs. 1.05 for speech-in-noise condition; $F(1, 16) = 12.59$; $p < .003$) and their interaction ($F(1, 16) = 4.76$; $p < .05$). As in the previous analyses, post-hoc Sheffe's tests revealed a significant decrease of activation when noise was added in the comprehension task (p < .004), but not in the perceptual task (p = .56).

However, using a predefined statistical threshold (voxelwise p < .001 uncorrected) at the individual level to define the subject-specific ROIs may lead to variable ROI size across subjects and thus bias the results. Intra-subject consistency has even been questioned when highly selective thresholds were applied (Duncan et al., 2009). To address this issue, we performed an additional analysis in which the subject-specific vOT corresponded to the 50 most activated voxels (unthresholded) in the search

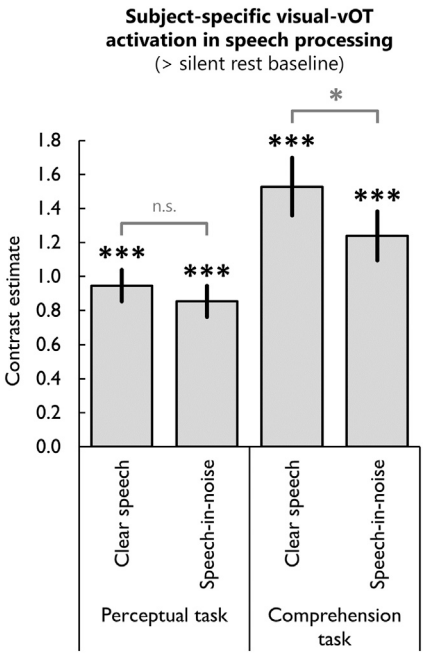


Fig. 9. Average contrast estimates for the four speech processing conditions against silent rest baseline, in the individual ROIs functionally defined based on the visual-VOT localizer (a voxelwise $p < .001$ uncorrected threshold). Error bars represent SEM. Black ***: corrected $p < .001$ obtained in the one-sample t-tests. Grey *: $p < .05$ for pairwise comparisons in the ANOVA's post-hoc test.

volume. Both one sample T-tests and ANOVA replicated the initial findings: Significant activations (all ps $< .001$) were observed in the four conditions. The main effects of task ($F(1, 20) = 14.19$; $p < .001$), presence of noise ($F(1, 20) = 8.55$; $p < .009$), and the interaction between the two factors ($F(1, 20) = 12.17$; $p < .003$) were significant.

2) The degree of overlap between "visual-vOT" and "auditory-vOT"

To assess the degree of overlap, at the individual level, between the vOT voxels activated during visual word processing and those activated during spoken sentences processing, a common search volume for all participants was first defined. It corresponded to the union of the group-level vOT cluster obtained in the visual words > consonant strings contrast (275 voxels, $p < .001$ unc.) and the group-level vOT cluster obtained in all speech processing tasks > silent rest contrast (214 voxels, $p < .001$ unc.). The resulting search volume contained 309 voxels (the voxels that were present in the cerebellum, according to the AAL2 atlas, were removed). Within this search volume, we identified, for each participant, the voxels that were activated (at the $p < .001$ uncorrected threshold) in the visual words > consonant strings contrast and those activated in the all speech processing tasks > silent rest contrast. Here again, the four participants who did not show any significant voxel in the visual contrast were excluded. The average numbers of voxels in these subject-specific functional ROIs were 51 (SD = 42) and 137 (SD = 65) voxels, respectively. Based on these ROIs, we found that, on average, 45 voxels per subject were activated in both language modalities. In terms of degree of overlap, this reflects that, on average, 83% of "visual-vOT" voxels were significantly activated in the speech processing contrast (SD = 28%) and 31% of "auditory-vOT" voxels were significantly activated in the visual word processing contrast (SD = 22%).

As in the first set of analyses, we also re-defined the individual ROIs by selecting the 50 most activated voxels (unthresholded) in the search volume in each of the two language modalities. With this criterion, the overlap ratio between these two ROIs was 52% on average (SD = 21%).

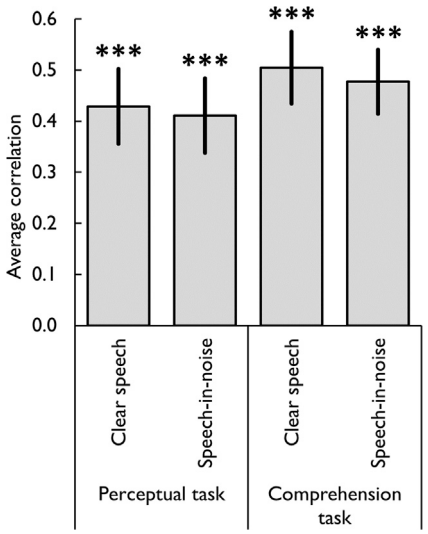


Fig. 10. Average correlations between the activations patterns induced by visual words and each of the four spoken sentence conditions. Error bars represent SEM. ***: $p < .0001$ (permutation test).

3) Correlations of activation patterns in spoken sentence processing and visual word processing

As a final test for the anatomo-functional convergence between vOT voxels that respond to written and spoken language inputs, we assessed the degree of within-subject similarity between the activation patterns obtained during visual word processing and each of the four speech processing conditions. We used the left vOT volume containing 309 voxels defined in the previous analysis. For each subject, we computed the Pearson's correlations between the activation value of the 309 voxels of the volume obtained during visual word processing (visual words > consonant strings) and each of the four speech processing conditions (contrasted against silent rest) (see Supplementary Fig. S2 for an illustration of this analysis in one representative participant). The significance of the correlations was assessed using a non-parametric permutation test with 5000 permutations: Subject labels were permuted, thus testing whether within-subject correlation was stronger than a distribution of across-subjects correlations. A highly significant positive correlation was found between the visual task and the comprehension task performed on clear speech (average $r = 0.50$; $p < .0001$), the comprehension task performed on speech-in-noise (average $r = 0.48$; $p < .0001$), the perceptual task performed on clear speech (average $r = 0.43$; $p < .0001$), and the perceptual task performed on speech-in-noise (average $r = 0.41$; $p < .0001$) (see Fig. 10).[1] Note that a few participants showed very weak correlations; median correlations were consequently higher than the group averages (i.e., median r were, respectively for the four above-mentioned conditions, .60, .59, 0.51 and 0.45). A repeated-measures ANOVA comparing the correlations coefficients obtained in the four analyses revealed neither main effects of task, of the presence of noise nor their interaction (all p. > 0.11). These results provide further support for an anatomo-functional correspondence of the vOT responses induced by spoken and written language inputs. They also indicate that this correspondence was stable and, thus, not sensitive to either bottom-up stimulus-driven or top-down task-

[1] To further ensure that these highly significant correlations are specific to the contrasts of interest rather than reflecting non-specific intra-subject correlations of image intensity across voxels, an additional analysis was also performed using a non-linguistic control contrast: noise-only rest vs. silent rest. As illustrated in Supplementary Fig. S2, no correlation was found between the activation obtained in the visual processing task and that obtained in the control contrast (average $r = -.03$; $p = .96$).

driven factor.

4. Discussion

Inspired by studies that reported the involvement of the left vOT in spoken word processing tasks (Yoncheva et al., 2010; Ludersdorfer et al., 2013, 2015, 2016), the current study further examined whether this key area of the reading network is also recruited during a more natural situation like spoken sentence processing. We specifically tested whether vOT responses to spoken sentences depend on two factors that contribute to task difficulty, i.e., task demands (low-level perception vs. comprehension task) and quality of spoken input (presence vs. absence of noises in speech signal). Using both whole brain and functionally defined subject-specific ROI approaches, our findings indicated that attending to spoken sentences systematically induced left vOT activation. However, the strength of the activation was modulated by both task demands and quality of the input. While the activation was generally enhanced in the comprehension compared to the perception task, the presence of noise that further increased task difficulty did not lead to an increase of activity. On the contrary, the vOT activity was reduced, especially in the comprehension task.

Top-down activation of the left vOT by speech input is generally regarded as optional, occurring only in specific experimental conditions, for instance, when participants were required to process single words in relatively challenging tasks. So far, this activation has mainly been interpreted as reflecting a retrieval of word spellings (Booth et al., 2002; Cao et al., 2010; Dehaene and Cohen, 2011). A previous observation by Dehaene et al. (2010) showed vOT activation during an auditory lexical decision task but not during passive listening of sentences which provided an argument in favor of this view.

Using speech processing tasks in which participants were required to process spoken sentences in different contexts, our findings do not support the claim that vOT activation is restricted to single word processing. On the contrary, the area appears as part of the speech processing network. As shown in the analysis of the global activation pattern elicited by spoken sentences compared to silent baseline (cf. Figs. 3 and 7), there is a wide spread activity that includes areas in the temporal cortices typically involved in spoken language processing as well as those that are parts of the semantic, articulatory and also orthographic systems. Interestingly, the activity within the left vOT was present even in the low-level perceptual task that did not require an analysis of linguistic content of speech inputs. However, as will be discussed further below, in this specific task, the whole brain analysis showed that the vOT activity only reached significance when speech were clearly presented but not when it was embedded in noises.

Before discussing the prominent roles of task demands and signal quality on the occurrence and amplitude of vOT activation, it is worth noting that the present findings also address a methodological issue raised by previous studies, that is, the impact of the baseline condition on the significance level of the cross-modal activity within the left vOT (Ludersdorfer et al., 2013, 2015; Yoncheva et al., 2010). Although, unlike some previous studies (e.g., Yoncheva et al., 2010), no deactivation of vOT was observed in any spoken sentence processing conditions, we clearly observed an enhancement of vOT responses when the activity in these active conditions was contrasted against the MSB noise baseline rather than against the typical silent baseline. This "artificial" enhancement was explained by the reduction of activity in the extrastriate areas during auditory processing of MSB noise (cf. Fig. 7C) which corresponds to the "cross-modal sensory suppression" phenomenon (Laurienti et al., 2002).

4.1. The role of task demands on the strength of vOT activation

The nature of the task was one of the main factors that affected the strength of vOT activation during speech processing. This was attested by the presence of activity in this area and in the left inferior frontal regions associated with phonological (pars opercularis) or high-level semantic processing (pars orbitalis) (see Poldrack et al., 1999), in the comprehension task > perception task contrast. This result confirms previous observations that the involvement of the vOT in speech processing is strengthen when participants' attention was drawn upon linguistic contents of speech stimuli (Yoncheva et al., 2010). However, the increase contribution of this area in the comprehension compared to the perceptual task may appear surprising considering that the literature on semantic processing classically reports the involvement of the inferior frontal, middle temporal or parietal inferior regions (see Binder et al., 2009). In their review article, Hickok and Poeppel (2007) considered the posterior inferior temporal lobe to be part of the" ventral stream" that is involved in the conceptual aspects of speech processing. Rodd et al. (2005) also reported an increase of activity in the left posterior inferior temporal gyrus, together with the left inferior frontal gyrus when comparing brain activity induced by high-ambiguity spoken sentences to that induced by low-ambiguity sentences. However, the posterior inferior temporal lobe mentioned in these studies seems to be located more anteriorly and laterally to the classic location of the visual-word sensitive vOT, thus perhaps corresponding to the lateral inferotemporal multimodal area reported by Cohen et al. (2004).

The literature on the functional role of the left vOT does not support the explanation that the area is specifically involved in a conceptual semantic function (e.g., Dehaene and Cohen, 2011). An alternative explanation of its increase contribution in the comprehension task could be that the activation level within the area depends on the processing load required by language tasks. The Local Combination Detector model proposed by Dehaene et al. (2005), and supported by fMRI results (Vinckier et al., 2007), explains the emergence of a visual-word sensitive area by proposing an existence of hierarchically organized occipito-temporal ventral pathway, composed of neurons tuned to the detection of the simplest visual forms in the posterior portion of the pathway, up to complete letters and full words in its anterior portion. However, this model that focuses on feedforward connections may have neglected the influence of top-down mechanisms underpinning the functional connections of the area with other linguistic regions on the emergence of this posterior-to-anterior gradient or on the location of the Visual Word Form Area itself. Using similar stimuli as Vinckier et al. (2007), Levy et al. (2008), replicated their finding but attributed the posterior-to-anterior recruitment of the ventral temporal pathway to an increase in stimuli's "linguistic processing load" rather than to their visual complexity alone. The authors also reported that the activity in other regions, notably the left IFG, was also sensitive to this hierarchy. An influence of non-visual linguistic demands of vOT activation was also shown Twomey et al. (2011), who found stronger top-down activation when the stimuli or the task placed increased demands on phonological processing.

The idea that processing load or task difficulty is a critical factor that drives the spread of neural activity from the primary sensory cortices to remote brain areas is also compatible with the hierarchical organization of speech processing, according to which, regions more and more distant from the auditory cortex are recruited for higher level processing (Davis and Johnsrude, 2003; Scott and Johnsrude, 2003). However, as will be described below, the results obtained in the conditions where speech signals were degraded by the presence of background noise indicate that processing load or task difficulty alone cannot entirely account for the activation pattern of the vOT.

4.2. The role of speech signal quality on the strength of vOT activation

As expected, degradation of the speech signal clearly affected task performance, confirming that this manipulation made speech processing more difficult. The deleterious effect of background noise was particularly pronounced in the comprehension task where correct detection of false statements dropped from 88% to 63%. However, the brain activation pattern associated with this manipulation is more complex. While the presence of noises mainly induced higher activity in the left-

dominant STG (Davis et al., 2011; Obleser et al., 2007; Wild et al., 2012) and the inferior and middle frontal gyri in the right hemisphere, it led to a decrease of activity in several areas in the left-hemisphere such as the middle occipital/angular gyrus and the posterior middle temporal gyrus that are involved in storage and retrieval of semantic knowledge (Humphries et al., 2007; Pallier et al., 2011).

Both whole brain and subject-specific fROI analyses showed that adding noise into speech signals decreases the activity within the left-vOT. Coherently with the behavioral data, the reduction is strongest in the comprehension task. The whole brain analysis also showed the same activation pattern in the anterior portion of Broca's area (pars triangularis) considered to be a convergence zone of written and spoken word processing stream (Liuzzi et al., 2017; Montant et al., 2011) and bilateral precentral and postcentral gyri which are part of the articulatory system (Pulvermüller et al., 2006). The observation of a reduced brain activity contradicts the hypothesis that explains the increase of top-down vOT activation by task difficulty. In addition to the critical role of task demands, here, we further argue that this top-down activation is also modified by the intelligibility of speech signal. Since the linguistic demands of the comprehension task remained constant in both clear speech and speech-in-noise conditions, our finding indicated that a factor that reduced the SNR also reduced a spread of neural activity from the primary sensory cortices to higher-order areas of the language network. A similar observation was also reported in our previous study (Pattamadilok et al., 2017). Using written words as inputs, we showed that the activity within the semantic and phonological areas decreased when the visibility of the written words decreased. Furthermore, the strength of neural propagation to remote areas further interacted with task demands such that the activity within the areas that are task-relevant was more strongly correlated with the visibility of the inputs than the activity within the areas that are not task-relevant.

Here, the presence of noise that increased acoustic and phonological processing load in the STG resulted in a disengagement of top-down vOT activation. The fact that the noise effect was restricted to the comprehension task may reflect a redistribution of limited cognitive resources, already mobilized by the demanding semantic task, towards the processing of the acoustic-phonetic features of the degraded input. Such redistribution of resources to compensate for reduced speech intelligibility has already been reported in the past. For instance, Vagharchakian et al. (2012) showed that the activation in high-level linguistic regions such as the IFG and the STS collapsed when the degradation of the speech signal reached a critical point, while the activation in sensory areas still increased linearly with the degree of speech degradation. Through a joint manipulation of semantic predictability and spectral degradation, Obleser et al. (2007) also observed a set of areas (angular gyrus, lateral prefrontal cortex, posterior cingulate cortex) whose activity was modulated by an interaction between these two factors: The high predictability condition yielded stronger activity than the low predictability condition only when spoken sentences were presented at an intermediate level of degradation, but not when they were not degraded or strongly degraded. This finding suggested that these brain areas were involved when needed and only if speech comprehension still succeeds despite adverse acoustic conditions. The activity of the left vOT observed in our study showed a similar pattern. It most strongly contributes to speech comprehension when this could be achieved at a reasonable accuracy level.

In sum, the opposing effects of task demands and quality of speech signal indicates that processing difficulty is not the only underlying factor that determines the strength of the vOT activation in response to speech. As previously argued (Pattamadilok et al., 2017, Price et al., 1997 and Price and Devlin, 2011), one may make a distinction between "automatic" and "strategic" propagations of neural activity from the primary sensory cortices to higher-order brain regions. While the automatic propagation of brain activity is mainly driven by the strength or SNR of the sensory input, this passive propagation could be modulated by a strategic process driven by attention and task demands. In the case of the vOT contribution to speech processing reported here, the interplay between these two mechanisms results in a graded activation pattern illustrated in Fig. 7A, ranging from the lowest activation in the perception task performed on low SNR speech signals to the highest activation in the comprehension task performed on high SNR speech signals. However, the allocation of the cognitive resources to the bottom-up and top-down factors seems to rely on a complex mechanism that involves several brain areas, especially in the frontal cortex whose activation was affected by the interaction between task demands and signal quality (cf., Table S3) (see also Awh et al., 2012). More extensive analyses taking into account the temporal dynamics of the connections between these areas, and the role they play in the communication between the auditory and visual systems, is needed to draw a complete picture of the phenomenon.

A unique region within the ventral occipitotemporal pathway is involved in the processing of written words and spoken sentences.

So far, most of the studies that reported left vOT activation in non-orthographic spoken word processing tasks have relied on literature-based coordinates to make the assumption that the area in the ventral occipitotemporal pathway that responds to spoken inputs is in the same location as the one recruited during written word processing (e.g., Ludersdorfer et al., 2016; Yoncheva et al., 2010). Nevertheless, as stated in the Introduction, due to the possible existence of distinct unimodal and multimodal areas in the vOT region (Cohen et al., 2004), and more generally to inter-subject heterogeneity in activation and anatomical characteristics, additional precautions must be taken before assuming an anatomical correspondence between functional maps obtained in different fMRI sessions or studies.

One approach that we used to examine the similarity between the responses induced by written and spoken inputs is to compare the activation patterns of sub-regions within the left ventral visual pathway in response to the two types of stimulus. To this aim, the activation estimates (betas) induced by spoken sentences and written words were extracted from six 6-mm radius spherical ROIs created from the coordinates from Vinckier et al.' s study (2007). As illustrated in Fig. 8, the two language modalities showed the same posterior-to-anterior gradient, both peaking around $y = -56$ and -48, corresponding to the location of the VFWA (Cohen et al., 2002). This observation provides a piece of evidence in favor of a homology between the area that responded to spoken inputs and the VWFA described in the literature. In addition to this global picture, a more precise anatomical correspondence between the vOT voxels activated by spoken and written language inputs was examined at the individual subject level. We applied the subject-specific fROI approach that consists in comparing, within each participant, the vOT voxels activated by spoken inputs to those activated by written inputs. These more fine-grained analyses showed that part of the voxels activated by written words was also activated by spoken sentences. Interestingly, these voxels also showed an activation profile similar to that found in the whole brain group analysis, in terms of the sensibility to task demands and signal quality manipulations. Overall, it is rather surprising that despite the differences between the written and spoken inputs (single words briefly flashed on the screen vs. full sentences unfolding in time) and the experimental paradigms that were used to extract of the activation maps of the two language modalities, the degree of overlap between the vOT voxels activated by the written and spoken inputs is still above 50% (when the 50 most activated voxels for each language modality were considered). The similarity between their activation patterns was further demonstrated in the analysis of the spatial pattern of neural responses in the left vOT, which is less sensitive to the variations in the overall activation extent and intensity between the two experiments. The analysis showed highly significant correlations between the activations obtained during written word processing and in each of the four speech processing situations. Contrary to the activation intensity that varies with task demands and signal quality, these correlations are statistically equivalent in all experimental conditions (i.e., coefficient between 0.50 and 0.41). It is indicative that although there is a variation in the activation level across conditions, the anatomo-functional correspondence of the vOT responses to spoken and

written language inputs remains stable. This conclusion is consistent with the idea that a same set of voxels responds to both spoken and written words in a given individual although the strength of their responses varies with task demands and quality of speech signal.

4.3. The nature of information encoded in the left vOT

The subject-specific ROI analyses suggest that a significant part of the vOT voxels that responded to written words also responded spoken sentences. So far, the literature mainly explains this top-down activation by two mechanisms. According to the '*orthographic tuning hypothesis*' proposed by Dehaene and colleagues (Cohen et al., 2004; Dehaene and Cohen, 2011; Dehaene et al., 2005), the left-vOT contains neurons are selectively tuned to written language input. These orthographic coding neurons could nevertheless be activated in a top-down fashion by a spoken input once it has been converted into an orthographic code. This hypothesis has provided a plausible explanation to what happens during single word processing, especially in difficult tasks such as lexical decision, spelling or meta-phonological tasks, where the activation of non-phonological codes including orthography has been consistently reported (Lafontaine et al., 2012; Seidenberg and Tanenhaus, 1979; Ziegler and Ferrand, 1998). However, it is unclear how this online conversion of phonological to orthographic representation could explain the present finding where spoken inputs are full sentences. Although one could rightly argue that the sentences might be partially converted in their spelling forms and activated the orthography-encoding neurons in the left vOT, the possibility that these incomplete orthographic representations matched the written words presented during the "Visual-vOT" localizer session and, thus, led to more than 50% overlapping activation of the two language modalities seems unlikely.

An alternative mechanism proposed by Price and Devlin (2011), who assumed that neuronal populations in vOT are not selectively tuned to orthographic inputs, seems to provide a more flexible mechanism to account for the present finding. According to their '*Interactive Account*', vOT is an interface between bottom-up sensory inputs and top-down predictions that are generated based on prior knowledge on the association of the visual input with phonological and semantic representations. The same neuronal populations in the area can contribute to different functions depending on the regions with which they interact and the processing context. As a result, the orthographic, semantic and phonological information is processed within distributed but interconnected neural networks.

Within this framework, spoken sentences would be processed within these large networks that encompass the left vOT, without assuming that the spoken inputs need to be converted into written forms. Although we do not have direct evidence supporting this interpretation, some recent findings provide arguments supporting the idea that this area within the ventral pathway may not only encode orthographic representations. Using representational similarity analyses, Zhao et al. (2016) observed a significant association between the pattern of phonological similarity of written Chinese words and neural responses in the anterior and middle fusiform gyrus, thus, suggesting that neurons within these areas also represent phonology. A similar conclusion was obtained in our recent study where transcranial magnetic stimulation was combined with an adaptation protocol to examine the properties of left-vOT neurons. Our findings suggest a co-existence of functionally segregated neuronal populations that selectively respond to written or to spoken language modality (Pattamadilok et al., 2019). Accordingly, the vOT responses to spoken sentences reported here could at least partly reflect the activity of neurons encoding spoken language that are intermingled with those encoding written language. Further brain-imaging studies using techniques that provide a higher spatial resolution are needed to demonstrate such spatial segregation between the two types of neuronal populations (e.g., Gentile et al., 2017).

In conclusion, the current study confirms that vOT responses to speech is not restricted to single word processing or to difficult tasks but also generalizes to more natural processing situations like sentence perception and comprehension. This finding could not entirely be explained by the simple conversion of spoken sentences into orthographic representations. It raises the possibility that neurons within this part of the ventral pathway might also encode spoken inputs. However, the present study only focuses on detailed analyses of the neural responses within a specific part of the ventral occipital pathway that corresponds to the location of the Visual Word Form Area. As mentioned earlier, the activity within this area as well as its sensitivity to bottom-up and top-down factors could be driven by a complex mechanism that involves several brain areas within and outside the language network. In line with this remark, Ludersdorfer et al., 2019 recently examined the effects of task demands and stimuli on the activation and connectivity in superior and inferior parts of the left vOT. Their dynamic causal modeling provides an insightful observation that the superior and inferior vOTs differentially drive the activation in the anterior portion of the occipito-temporal sulcus, whose MNI coordinates corresponded to the location of the vOT in the present study. This observation combined with our finding that the activation of the vOT co-occurred with the activation of many areas in the frontal cortex (cf. Table S3) motivate more global investigations that consider the left vOT as part of an extended dynamic neural network.

Acknowledgments

This work was supported by the French Ministry of Research: ANR-13-JSH2-0002 (to C.P.), ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). It was performed in the Center IRM-INT (UMR 7289, AMU-CNRS), platform member of France Life Imaging network (grant ANR-11-INBS-0006).We warmly thank Dr. Agnès Trébuchon for taking medical responsibility during the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116135.

References

Adank, P., Devlin, J.T., 2010. On-line plasticity in spoken sentence comprehension: adapting to time-compressed speech. Neuroimage 49, 1124–1132.
Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B., Zilles, K., 1999. Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412, 319–341.
Awh, E., Belopolsky, A.V., Theeuwes, J., 2012. Top-down versus bottom-up attentional control: a failed theoretical dichotomy. Trends Cogn. Sci. 16, 437–443.
Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45, 2883–2901.
Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebr. Cortex 19, 2767–2796.
Bolger, D.J., Perfetti, C.A., Schneider, W., 2005. Cross-cultural effect on the brain revisited: universal structures plus writing system variation. Hum. Brain Mapp. 25, 92–104.
Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M., 2003. Relation between brain activation and lexical performance. Hum. Brain Mapp. 19, 155–169.
Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2002. Functional anatomy of intra-and cross-modal lexical tasks. Neuroimage 16, 7–22.
Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2004. Development of brain mechanisms for processing orthographic and phonologic representations. J. Cogn. Neurosci. 16, 1234–1249.
Brem, S., Bach, S., Kucian, K., Kujala, J.V., Guttorm, T.K., Martin, E., Lyytinen, H., Brandeis, D., Richardson, U., 2010. Brain sensitivity to print emerges when children learn letter–speech sound correspondences. Proc. Natl. Acad. Sci. 107, 7939–7944.
Brett, M., Johnsrude, I.S., Owen, A.M., 2002. The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3, 243.
Büchel, C., Price, C., Friston, K., 1998. A multimodal language region in the ventral visual pathway. Nature 394, 274.
Burianova, H., Grady, C.L., 2007. Common and unique neural activations in autobiographical, episodic, and semantic retrieval. J. Cogn. Neurosci. 19, 1520–1534.

Burton, M.W., LoCasto, P.C., Krebs-Noble, D., Gullapalli, R.P., 2005. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. Neuroimage 26, 647–661.
Cao, F., Khalid, K., Zaveri, R., Bolger, D.J., Bitan, T., Booth, J.R., 2010. Neural correlates of priming effects in children during spoken word processing with orthographic demands. Brain Lang. 114, 80–89.
Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., Michel, F., 2000. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123 (Pt 2), 291–307.
Cohen, L., Jobert, A., Le Bihan, D., Dehaene, S., 2004. Distinct unimodal and multimodal regions for word processing in the left temporal cortex. Neuroimage 23, 1256–1270.
Cohen, L., Lehéricy, 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.
Cone, N.E., Burman, D.D., Bitan, T., Bolger, D.J., Booth, J.R., 2008. Developmental changes in brain regions involved in phonological and orthographic processing during spoken language processing. Neuroimage 41, 623–635.
Davis, M.H., Ford, M.A., Kherif, F., Johnsrude, I.S., 2011. Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. J. Cogn. Neurosci. 23, 3914–3932.
Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431.
Dehaene-Lambertz, G., Monzalvo, K., Dehaene, S., 2018. The emergence of the visual word form: longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLoS Biol. 16, e2004103.
Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15, 254–262.
Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. Trends Cogn. Sci. 9, 335–341.
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.
Desroches, A.S., Cone, N.E., Bolger, D.J., Bitan, T., Burman, D.D., Booth, J.R., 2010. Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. Brain Res. 1356, 73–84.
Duncan, K.J., Pattamadilok, C., Knierim, I., Devlin, J.T., 2009. Consistency and variability in functional localisers. Neuroimage 46, 1018–1026.
Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., Kanwisher, N., 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. J. Neurophysiol. 104, 1177–1194.
Fiebach, C.J., Friederici, A.D., 2004. Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. Neuropsychologia 42, 62–70.
Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., Zilles, K., 2007. Cortical folding patterns and predicting cytoarchitecture. Cerebr. Cortex 18, 1973–1980.
Gentile, F., van Atteveldt, N., De Martino, F., Goebel, R., 2017. Approaching the ground truth: revealing the functional organization of human multisensory STC using ultra-high field fMRI. J. Neurosci. 37, 10104–10113.
Grainger, J., Ziegler, J.C., 2007. Cross-code Consistency in a Functional Architecture for Word Recognition. Single-word Reading. Psychology Press, pp. 142–170.
Harm, M.W., Seidenberg, M.S., 1999. Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol. Rev. 106, 491.
Harm, M.W., Seidenberg, M.S., 2004. Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychol. Rev. 111, 662.
Henson, R.N., 2015. Design efficiency. In: Toga, A.W. (Ed.), Brain Mapping: an Encyclopedic Reference. Academic Press: Elsevier, pp. 489–494.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2007. Time course of semantic processes during sentence comprehension: an fMRI study. Neuroimage 36, 924–932.
Jobard, G., Crivello, F., Tzourio-Mazoyer, N., 2003. Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. Neuroimage 20, 693–712.
Juch, H., Zimine, I., Seghier, M.L., Lazeyras, F., Fasel, J.H., 2005. Anatomical variability of the lateral frontal lobe surface: implication for intersubject variability in language neuroimaging. Neuroimage 24, 504–514.
Lafontaine, H., Chetail, F., Colin, C., Kolinsky, R., Pattamadilok, C., 2012. Role and activation time course of phonological and orthographic information during phoneme judgments. Neuropsychologia 50, 2897–2906.
Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.-F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14, 420–429.
Levy, J., Pernet, C., Treserras, S., Boulanouar, K., Berry, I., Aubry, F., Demonet, J.-F., Celsis, P., 2008. Piecemeal recruitment of left-lateralized brain areas during reading: a spatio-functional account. Neuroimage 43, 581–591.
Liberman, A.M., Whalen, D.H., 2000. On the relation of speech to language. Trends Cogn. Sci. 4, 187–196.
Liuzzi, A.G., Bruffaerts, R., Peeters, R., Adamczuk, K., Keuleers, E., De Deyne, S., Storms, G., Dupont, P., Vandenberghe, R., 2017. Cross-modal representation of spoken and written word meaning in left pars triangularis. Neuroimage 150, 292–307.
Ludersdorfer, P., Kronbichler, M., Wimmer, H., 2015. Accessing orthographic representations from speech: the role of left ventral occipitotemporal cortex in spelling. Hum. Brain Mapp. 36, 1393–1406.
Ludersdorfer, P., Price, C.J., Duncan, K.J.K., DeDuck, K., Neufeld, N.H., Seghier, M.L., 2019. Dissociating the functions of superior and inferior parts of the left ventral occipito-temporal cortex during visual word and object processing. NeuroImage.
Ludersdorfer, P., Schurz, M., Richlan, F., Kronbichler, M., Wimmer, H., 2013. Opposite effects of visual and auditory word-likeness on activity in the visual word form area. Front. Hum. Neurosci. 7, 491.
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.
McCandliss, B.D., Cohen, L., Dehaene, S., 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn. Sci. 7, 293–299.
Montant, M., Schön, D., Anton, J.-L., Ziegler, J.C., 2011. Orthographic contamination of Broca's area. Front. Psychol. 2, 378.
New, B., Pallier, C., Brysbaert, M., Ferrand, L., 2004. Lexique 2: a new French lexical database. Behav. Res. Methods Instrum. Comput. 36, 516–524.
Obleser, J., Kotz, S.A., 2010. Expectancy constraints in degraded speech modulate the language comprehension network. Cerebr. Cortex 20, 633–640.
Obleser, J., Wise, R.J., Dresner, M.A., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J. Neurosci. 27, 2283–2289.
Pallier, C., Devauchelle, A.-D., Dehaene, S., 2011. Cortical representation of the constituent structure of sentences. Proc. Natl. Acad. Sci. 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.
Pattamadilok, C., Planton, S., Bonnard, M., 2019. spoken language coding neurons in the visual word form area: evidence from a TMS adaptation paradigm. Neuroimage 186, 278–285.
Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage 10, 15–35.
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 Cogn. Sci. 15, 246–253.
Price, C.J., Moore, C.J., Humphreys, G.W., Wise, R.J., 1997. Segregating semantic from phonological processes during reading. Journal of Cognitive. Neuroscience 9, 727–733.
Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F.M., Hauk, O., Shtyrov, Y., 2006. Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. 103, 7865–7870.
Reich, L., Szwed, M., Cohen, L., Amedi, A., 2011. A ventral visual stream reading center independent of visual experience. Curr. Biol. 21, 363–368.
Rodd, J.M., Davis, M.H., Johnsrude, I.S., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cerebr. Cortex 15, 1261–1269.
Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage 30, 1088–1096.
Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107.
Seghier, M.L., 2013. The angular gyrus: multiple functions and multiple subdivisions. The Neuroscientist 19, 43–61.
Seidenberg, M.S., Tanenhaus, M.K., 1979. Orthographic effects on rhyme monitoring. J. Exp. Psychol. Hum. Learn. Mem. 5, 546.
Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage 16, 765–780.
Twomey, T., Duncan, K.J.K., Price, C.J., Devlin, J.T., 2011. Top-down modulation of ventral occipito-temporal responses during visual word recognition. Neuroimage 55, 1242–1251.
Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., Dehaene, S., 2012. A temporal bottleneck in the language comprehension network. J. Neurosci. 32, 9089–9102.
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.
Whittingstall, K., Bernier, M., Houde, J.-C., Fortin, D., Descoteaux, M., 2014. Structural network underlying visuospatial imagery in humans. Cortex 56, 85–98.
Wild, C.J., Davis, M.H., Johnsrude, I.S., 2012. Human auditory cortex is sensitive to the perceived clarity of speech. Neuroimage 60, 1490–1502.
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. Cerebr. Cortex 20, 622–632.
Zhao, L., Chen, C., Shao, L., Wang, Y., Xiao, X., Chen, C., Yang, J., Zevin, J., Xue, G., 2016. Orthographic and phonological representations in the fusiform cortex. Cerebr. Cortex 27, 5197–5210.
Ziegler, J.C., Ferrand, L., 1998. Orthography shapes the perception of speech: the consistency effect in auditory word recognition. Psychon. Bull. Rev. 5, 683–689.