

The visual word form area: a prelexical representation of visual words in the fusiform gyrus

Stanislas Dehaene,^{CA} Gurvan Le Clec'H, Jean-Baptiste Poline,¹ Denis Le Bihan¹ and Laurent Cohen²

Unité INSERM 334 and; ¹Unité de Neuro-Activation Fonctionnelle, IFR 49, Service Hospitalier Frédéric Joliot, CEA/DSV, 4 Place du Général Leclerc, 91401 Orsay; ²Institut de Neurologie, IFR 49, Hôpital de la Salpêtrière, AP-HP, Paris, France

Received 10 December 2001 accepted 3 January 2002

Event-related fMRI was used to test the hypothesis that the visual word form area in the left fusiform gyrus holds a modality-specific and prelexical representation of visual words. Subjects were engaged in a repetition-detection task on pairs of words or pronounceable pseudo-words that could be written or spoken. The visual word form area responded only to written stimuli, not to

spoken stimuli, independently of their semantic content. We propose that the occasional activation of the fusiform gyrus when listening to spoken words is due to the topdown recruitment of visual orthographic or object representations. *NeuroReport* 13:321–325 © 2002 Lippincott Williams & Wilkins.

Key words: Reading; Temporal lobe; Word recognition

INTRODUCTION

Whenever a skilled reader is presented with a printed word, brain activity can be observed in a well-delineated area of the left fusiform gyrus [1–4], which is consistently found at the same location in Talairach space (about $x=-42$, $Y=-57$, $z=-6$), with a standard deviation of only about 0.5 cm [1]. This activation generally falls on the mesial edge of or within the occipito-temporal sulcus, about halfway along the antero-posterior axis of the fusiform gyrus. Our working hypothesis is that the left fusiform gyrus contains a visual word form area (VWFA) which computes an invariant structural representation of the visual word as an ordered sequence of abstract letter identities invariant for size, font, and case [5,6].

In support of this hypothesis, the VWFA responds with the same intensity whether words are presented in the left or in the right hemifield [1] (location invariance), and show equal repetition priming within and across case (case invariance) [4]. The visual word form hypothesis also implies that this area is prelexical. This fits well with the observation that the electrical and magnetic waves systematically elicited by visual words over the left ventral occipito-temporal region, and thought to originate from the left fusiform gyrus, have a short latency of 150–200 ms [1,3,7,8].

The hypothesis also implies that the VWFA is a unimodal visual area and is not normally implicated in spoken word processing. Indeed, neuropsychological studies confirm that lesions of the left posterior fusiform gyrus cause a unimodal deficit of word reading with sparing of auditory word comprehension and production [9,10]. Surprisingly, how-

ever, several neuroimaging studies have reported fusiform activations to words presented in a non-visual modality [11–19]. Correspondingly, some researchers have proposed that the fusiform gyrus is engaged in higher-level language processing, for instance in the phonological output lexicon, or in lexical access from semantics [20].

The goal of the present paper is to firmly demonstrate the prelexical and modality-specific character of the VWFA. As part of an exploratory experiment on category-specific representations, we measured single-event fMRI activation to written and spoken verbal stimuli while subjects performed an identical same-different judgment task on pseudo-words and on words from five different semantic categories. This provided an opportunity to measure, in the same individuals, the response of the VWFA to the same stimuli presented visually or auditorily in the same task on randomly intermixed trials. We predicted that the VWFA area should respond exclusively to the written stimuli, not to the spoken stimuli, regardless of their semantic content.

MATERIALS AND METHODS

Subjects: Seven right-handed, neurological normal volunteers (five women), aged between 19 and 34 years (mean 23 years 10 months), were studied. This research project was approved by the regional ethical committee, and all subjects gave their written informed consent.

Stimuli: Nineteen French words were selected within each of five semantic categories: tools, body parts, action verbs, animals, and numbers. They ranged in visual length from

three to nine letters (mean 6.06), in spoken length from one to three syllables (mean 1.62), and in frequency from 0.18 to 3.17 \log_{10} parts per million (mean 1.42). Thirty-eight matching pseudo-words obeying the phonotactic and graphotactic constraints of the French language were also generated (mean letters 5.92; mean syllables 1.74).

Imaging parameters: Functional images sensitive to blood oxygen level dependent (BOLD) contrast were obtained on a 3 T magnet (Bruker) with a T2* weighted gradient echo-planar imaging sequence (TR = 2.5 s, TE = 40 ms, angle = 90°, field of view 3.75 × 3.75 mm, matrix = 64 × 64). Twenty-four 6 mm axial cuts, covering the whole brain and the top of the cerebellum, were obtained every 2.5 s. Trials were presented every 15 s, corresponding to six fMRI repetition cycles, but fMRI images were acquired only during the last five cycles. During the first cycle, excitation gradients were maintained, but read-out gradients were omitted, thus leaving a quasi-silent period of 2.5 s during which the stimuli were presented with improved auditory comprehensibility. A total of 145 brain volumes were acquired consecutively for each run (five images in each of 29 trials), for a total of 580 images per subject (four runs). The first five images of each run were not included in the analysis. High-resolution images (3D gradient, echo inversion-recovery sequence, TI = 700 ms, TR = 1600 ms, FOV = 192 × 256 × 256 mm³, matrix = 256 × 128 × 256, slice thickness = 1.2 mm) were also acquired for anatomical localisation.

Procedure: On each trial, two consecutive stimuli (words or pseudo-words) were presented, either both visually or both auditorily (Fig. 1). The subjects' task was to decide whether or not the same stimulus had been repeated twice, by pressing one of two randomly assigned hand-held response keys. They had to perform this same-different judgement while neglecting irrelevant variations in the format of presentation. On visual trials, the first stimulus was always presented in upper case (angular size 1.6–5.5°), while the second was presented in lower case and in a 40% smaller type (angular size 1–3.3°). On auditory trials, the two stimuli were spoken by two different male voices.

Following a training run of 10 trials, subjects took four runs with one initial training trial and 28 experimental trials consisting of exactly one presentation of each combination of stimulus category (five semantic categories of words, plus two repetitions of the pseudo-word category), modality (auditory or visual), and the repetition factor (same or different words). On each trial, 200 ms after the fMRI scanner noise stopped, the first stimulus of the pair was presented, followed after 1200 ms by the second word. The duration of the stimuli was 500 ms in the visual modality, and varied from 300 to 700 ms in the auditory modality.

Statistical analysis: The analysis, performed with Statistical Parametric Mapping version 99 (SPM99), involved correction for slice acquisition delays and subject motion, normalization to the Montreal Neurological Institute template, and spatial smoothing (5 mm). The experimental design was a full orthogonal design with six stimulus categories (names of tools, body parts, action verbs, animals, or numbers, and pseudo-words), two stimulus modalities

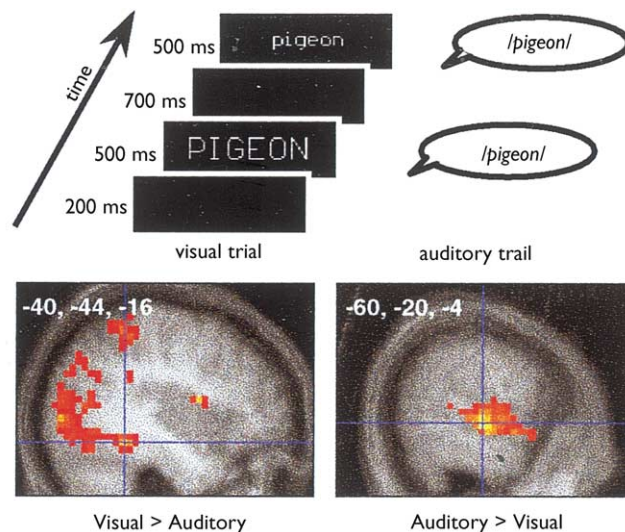


Fig. 1. Design of the experiment (top) and group results (bottom). Subjects saw or heard pairs of words belonging to the same semantic category, and decided whether the two words were identical, neglecting changes in font or speaker voice. Written words lead to increased activation throughout the occipito-temporal pathway, while spoken words activated the bilateral superior temporal gyrus.

(auditory or visual), and two levels of the word repetition factor (repeated or different stimuli). Activation on each of the corresponding 24 trials types was modelled by a linear combination of the standard SPM hemodynamic function and its temporal derivative. Four additional variables of non-interest modelled constant differences across the four runs. Long-term signal variations were eliminated with a high-pass filter set at 120 s. Low-pass filtering was achieved by convolution with the hemodynamic response function.

RESULTS

Behavior: The task was easily performed (visual stimuli: mean reaction time 757 ms, 0.2% errors; auditory stimuli, mean reaction time 863 ms, 0.8% errors). Responses were faster in the visual modality than in the auditory modality ($F(1,6) = 10.9$; $p = 0.016$). Response times to the six categories of targets also differed ($F(5,30) = 3.63$, $p = 0.011$), due to faster responding to words than to pseudo-words ($F(1,6) = 23.6$, $p = 0.003$), with no difference between the five categories of words ($F(4,24) = 2.32$, $p > 0.05$) and no interaction with modality.

Group analysis: A random-effects group analysis of the modality effects was performed (voxelwise $p < 0.01$, cluster-level $p < 0.05$ corrected). As predicted, greater activation for written than for spoken stimuli was observed in the entire occipito-temporal ventral stream, from occipital extrastriate cortex, where activation was bilateral (peak coordinates $-32, -88, -4$ and $28, -88, 12$) to the anterior fusiform gyrus where activation was bilateral but stronger on the left (Fig. 1; subpeaks at $-44, -76, -12$; $-52, -64, -8$; and $-40, -44, -16$). Activation was also observed in the left anterior cingulate, left and right posterior intraparietal regions, and the left anterior insula. Contrariwise, greater activation to

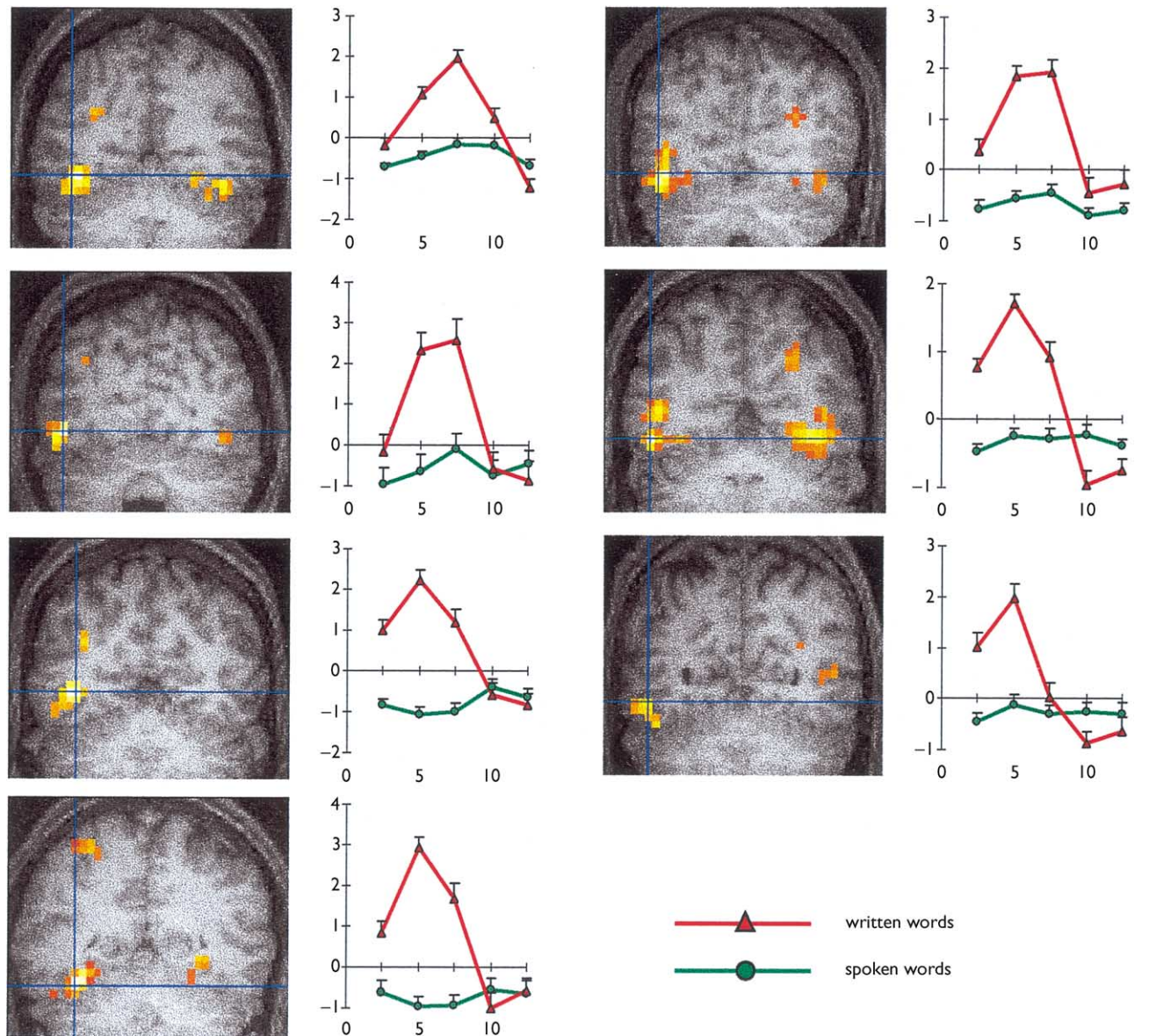


Fig. 2. Anatomical localization and modality-specificity of the visual word form area. In all seven subjects, an area of the left fusiform gyrus responded robustly to written words, but not to spoken words.

spoken stimuli was found in the left and right superior temporal gyri (Fig. 1; peaks at 60, -24, 8; -60, -20, -4). Importantly, examination of the main effect of activation to spoken words revealed no activation within the ventral occipito-temporal cortex, including the fusiform gyrus.

Single-subject analyses in the VWFA: Individual analyses (voxelwise $p < 0.01$, cluster-level $p < 0.05$ corrected) identified the VWFA as a cluster of voxels activated by visual words relative to rest in the left posterior fusiform gyrus at coordinates close to published estimates (mean (\pm s.d.) coordinates -41 ± 5 , -61 ± 7 , 8 ± 6 ; Fig. 2). Its anatomical localization was invariably on the banks of the left occipito-temporal sulcus that runs lateral to the fusiform gyrus.

Coronal slices revealed that at this level, the activation was highly asymmetrical in favour of the left hemisphere, though smaller activations were seen in the contralateral region of the right hemisphere in six of seven subjects.

We then performed individual analyses at the peak voxel, using the sensitive threshold of $p < 0.05$ uncorrected. None of the subjects showed any activation of the VWFA peak by spoken words, even at the sensitive threshold of $p < 0.05$. As shown in Fig. 2, the event-related activation curves showed a clear phasic activation with a signal change of about 2% when written stimuli were presented, but remained essentially flat when spoken stimuli were presented. Direct comparison revealed significantly larger activation for visual than for auditory trials in all subjects (all t values > 8 ; average $t = 11.3$).

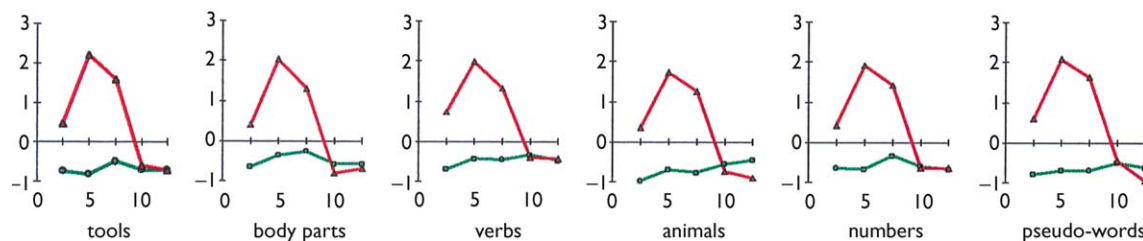


Fig. 3. Asemantic response profile of the visual word form area. Curves show the event-related response of the visual word form area averaged across subjects as a function of word category and modality (red = visual, green = auditory). Activation was independent of the semantic category of words and was very slightly enhanced when meaningless pseudo-words were presented.

The hypothesis that the VWFA is involved in prelexical visual word processing predicted that its response should be no greater to words than to pseudo-words. In fact, all subjects showed a small trend in the opposite direction of a greater activation to pseudo-words than to words (sign test, $p < 0.01$). However, the effect appeared quite small on activation curves (Fig. 3), and in individual analyses it was significant in only 2 of 7 subjects ($p = 0.043$ and $p = 0.002$). An F test was used to compare activations to the five semantic categories of words. None of the subjects showed any significant difference. The profile of VWFA activations was remarkably identical across semantic categories, although each subject only received eight trials within a given word category in the visual modality.

DISCUSSION

As reported previously [1,4], fMRI easily identified, in each individual subject, a strong left-lateralized activation to visual words on the lateral border of the fusiform gyrus. The results are compatible with the hypothesis that this area is involved in a visual and prelexical representation. In all subjects, its activity was found to be strictly visual, with no hint of an activation to spoken words.

Furthermore, no differential activation was found in the VWFA as a function of the semantic category of the stimuli. Finally, even meaningless pseudo-words evoked very strong activation in this area, confirming its involvement in prelexical processing. If anything, slightly more activity was observed when pronounceable pseudo-words were compared to real words of the same length. Note that subjects were significantly slower with pseudo-words than with words, perhaps because performance of the same-different task on words was supported in part by lexical and semantic levels of analysis not available for pseudo-words. After convolution with the hemodynamic response function, a slightly greater duration of fusiform activation for pseudo-words would translate into a slightly greater BOLD activation, as observed.

Our results suggest that the fusiform gyrus does not make a compulsory contribution to spoken word processing, confirming previous studies that have observed no activation of this region in a variety of language listening situations [19,21]. Why then did several other studies report fusiform activation by spoken words? As a working hypothesis, we propose that those fusiform activations reflect the top-down activation of either orthographic or

object knowledge in visual cortex, and are seen only with two types of tasks.

First, some tasks may require subjects to form a mental image of the written form of spoken words, thus calling for the activation of visual word representations within the VWFA. For instance, fusiform activations are observed when Japanese subjects are asked to write complex kanji ideograms or to imagine doing so [22], a task which presumably leads to explicit visual mental imagery of the words. The top-down recruitment of orthographic representations may also be implicit [11,12]. For instance, a covert activation of the VWFA can be expected in the phoneme-detection task with spoken words, because this task is known in psychology to be heavily influenced by word orthography.

Second, other tasks may promote the creation of visual mental images appropriate to the semantic contents of spoken words. Those tasks would not activate representations in the VWFA proper. Rather, they would lead to broad bilateral activations of visual object representations in the fusiform gyrus. In the literature, indeed, most of the studies that reported fusiform activations induced by spoken words used contrasts between semantic and non-semantic conditions, often with a special emphasis on mental imagery. For instance, the fusiform activation observed by Wise *et al.* was identified as an area whose activity was correlated with word imageability, irrespective of stimulus modality [15]. Similar activation foci were observed when contrasting the generation of mental images from spoken words relative to passive word listening [16], semantic *vs* phonological judgement tasks [17], story listening *vs* backward speech [18], attending to concrete words or meaningful sounds *vs* non-words or meaningless noises [19], semantic *vs* acoustic judgement tasks [11], or word-stem completion *vs* fixation [14]. We suggest that those activations are similar to those observed when subjects generate mental images of various categories of objects, in the same areas that are activated during actual perception [23].

CONCLUSION

The present experiment, together with similar past work [1,4] constrains the coding of words in the VWFA. This code is strictly visual, yet invariant for location, font and size. In the monkey inferotemporal cortex, visual neurons with wide receptive fields, selective to high-level visual features, and invariant for size and position of the visual stimuli have been recorded [24]. We tentatively propose that an analo-

gous combinatorial neural code, acquired as one learns to become an expert in reading, underlies visual orthographic processing in the human fusiform gyrus.

REFERENCES

1. Cohen L, Dehaene S, Naccache L *et al.* *Brain* **123**, 291–307 (2000).
2. Fiez JA and Petersen SE. *Proc Natl Acad Sci USA* **95**, 914–21 (1998).
3. Puce A, Allison T, Asgari M *et al.* *J Neurosci* **16**, 5205–5215 (1996).
4. Dehaene S, Naccache L, Cohen L *et al.* *Nature Neurosci* **4**, 752–758 (2001).
5. Warrington EK and Shallice T. *Brain* **103**, 99–112 (1980).
6. Hillis AE and Caramazza A. *Neurocase* **1**, 189–207 (1995).
7. Dehaene S. *NeuroReport* **6**, 2153–2157 (1995).
8. Salmelin R, Service E, Kiesilä P *et al.* *Ann Neurol* **40**, 157–162 (1996).
9. Binder JR and Mohr JP. *Brain* **115**, 1807–1826 (1992).
10. Leff AP, Crewes H, Plant GT *et al.* *Brain* **124**, 510–521 (2001).
11. Démonet J-F, Chollet F, Ramsay S *et al.* *Brain* **115**, 1753–1768 (1992).
12. Démonet J-F, Price C, Wise R *et al.* *Brain* **117**, 671–682 (1994).
13. Bochel C, Price C and Friston K. *Nature* **394**, 274–7 (1998).
14. Buckner RL, Koutstaal W, Schacter DL and Rosen BR. *Brain* **123**, 620–640 (2000).
15. Wise RJS, Howard D, Mummery CJ *et al.* *Neuropsychologia* **38**, 985–994 (2000).
16. D'Esposito M, Detre JA, Aguirre GK *et al.* *NeuroPsychologia* **35**, 725–30 (1997).
17. Chee MWL, O'Craven KM, Bergida R *et al.* *Hum Brain Mapp* **7**, 15–28 (1999).
18. Perani D, Paulesu E, Sebastian-Galles N *et al.* *Brain* **121**, 1841–1852 (1998).
19. Giraud AL and Price CJ. *J Cogn Neurosci* **13**, 754–765 (2001).
20. Brunswick N, McCrory E, Price CJ *et al.* *Brain* **122**, 1901–1917 (1999).
21. Binder JR, Frost JA, Hammeke TA *et al.* *Cerebr Cortex* **10**, 512–528 (2000).
22. Nakamura K, Honda M, Okada T *et al.* *Brain* **123**, 954–67 (2000).
23. Ishai A, Ungerleider LG and Haxby JV. *Neuron* **28**, 979–990 (2000).
24. Tanaka K. *Annu Rev Neurosci* **19**, 109–39 (1996).

Acknowledgements: This work was supported by a James S. McDonnell Centennial Fellowship.



LIPPINCOTT
WILLIAMS & WILKINS

Unauthorized Use
Prohibited