The Visual What For Area: Words and pictures in the left fusiform gyrus

Randi Starrfelt[a,*] and Christian Gerlach[b,c]

[a]Center for Visual Cognition, Department of Psychology, University of Copenhagen, Linnesgade 22, DK-1361 Copenhagen K, Denmark
[b]The Neurobiology Research Unit, N9201, and The PET and Cyclotron Unit, KF3982, Department of Clinical Physiology and Nuclear Medicine, Copenhagen University Hospital, Denmark
[c]Learning Lab Denmark, The Danish University of Education, Denmark

An area in the left fusiform gyrus labelled the Visual Word Form Area (VWFA) is claimed to be especially, or even selectively, responsive to words. We explored how stimulus type and task demands affect activity in this area by conducting a PET experiment where words and pictures were presented in two conditions that differed in demands on shape processing: colour decision and categorization. The subjects also performed an object decision task with pictures only. The imaging data revealed a main effect of stimulus type: rCBF was higher during word compared with picture processing. When compared individually for colour decision and categorization, the difference between words and pictures was only significant during colour decision, although a trend was present during categorization also. rCBF in the VWFA was highest during the object decision task, where only pictures were presented. Our findings indicate that the putative VWFA is activated more by written words than pictures, but only under certain circumstances. As demands on shape processing increase, the difference in activation between words and pictures decreases and can even be abolished. We suggest that activation in the VWFA could reflect shape configuration—the integration of shape elements into elaborate shape descriptions corresponding to whole objects or words. This process may be required to different degrees for pictures and words depending on task demands.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Reading; Visual recognition; Shape configuration; Fusiform gyrus; Functional imaging

Introduction

Recent functional imaging studies have identified a region in the middle part of the left fusiform gyrus which has been labelled the Visual Word Form Area (VWFA) (Cohen et al., 2000; McCandliss et al., 2003). It is claimed that this area is responsible for computing representations of "abstract letter identities invariant for parameters such as spatial position, size, font or case" (Cohen et al., 2003; p. 1314). This claim is based on two main lines of argument. First, this area is activated by words and letter-strings as compared to rest, fixation, or viewing of simple patterns like checkerboards (Cohen et al., 2000, 2002). Secondly, this area in the left hemisphere is activated regardless of hemifield of presentation in tachistochopic experiments, indicating that this activation represents processing of abstract, rather than stimulus specific, letter identities (Cohen et al., 2000). The VWFA has also been implicated in the disorder of pure alexia (Cohen et al., 2003; Gaillard et al., 2006), although this claim has recently been questioned (Hillis et al., 2005).

The existence of a cerebral area solely dedicated to processing of abstract letter or word forms, as well as the suggested name, has been challenged both on theoretical and empirical grounds (Price and Devlin, 2003, 2004), and it is still not clear which role this area might play in recognition of words and other visual stimuli. Depending on whether words, non-words, consonant strings or single letters are used, and what kind of stimuli they are compared to, the activation in the VWFA differs. The VWFA is quite consistently found activated when presentation of written words is contrasted with rest, fixation or simple visual patterns (Cohen et al., 2000, 2002), but this picture changes when letters and letter strings are contrasted with more complex stimuli, like pictures of faces (Puce et al., 1996) or objects (Joseph et al., 2003, 2006). For instance, in an fMRI study comparing processing of pictures of animals and single letters in passive viewing and silent naming tasks, Joseph et al. (2003) failed to find any fusiform area selectively activated by single letters. Also, Joseph et al. (2006) report that no area in the left fusiform gyrus is selectively activated by letters compared to pictures of objects when subjects perform a matching task. Single letter processing might not be a task that would activate the VWFA though, given that different parts of the fusiform gyrus are found activated when single letters are compared to letter strings (James et al., 2005, although see Joseph et al., 2006 for arguments to the contrary).

---

* Corresponding author. Fax: +45 3532 4802.
  E-mail address: randi.starrfelt@psy.ku.dk (R. Starrfelt).
  Available online on ScienceDirect (www.sciencedirect.com).

Regarding *what* kind of word representation is computed in the VWFA, the picture is also unclear: While some studies find this area more activated by real words than consonant strings or pseudowords (Cohen et al., 2002), others have found that activity in the VWFA increases as word frequency decreases, and is highest for pseudowords compared to real words (Kronbichler et al., 2004). The authors of the latter study claim that this effect of frequency on activation in the VWFA indicates that this area contributes to the recognition of visual words in a manner that supersedes the creation of abstract letter representations. The VWFA has also been implicated in non-visual processing, like semantic processing or phonological retrieval (Price et al., 2003; Vigneau et al., 2005). The focus of the present paper is limited to the suggested role of the VWFA in visual processing.

As noted by McCandliss et al. (2003 p.294) "there is ample evidence that object and face recognition can also activate this area to varying degrees", and, as pointed out by Price and Devlin (2004), studies are required to examine whether activation in this area is dissociable for words versus objects, and whether this might be dependent on task demands. Given the disparity of findings between studies comparing words to relatively simple visual stimuli versus studies comparing words to objects or faces, one could assume that activity in the VWFA is affected by visual complexity and demands on shape processing.

In order to investigate this hypothesis, we conducted a PET-study testing whether rCBF in the proposed visual word form area is greater for words than for pictures of objects, and whether this difference is modulated by the degree of shape analysis required to solve the task at hand. In the PET experiment, words and pictures were presented in two conditions that differed in demands on shape analysis: colour decision and categorization. During colour decision, subjects were to indicate whether stimuli were coloured white or yellow, a task requiring little shape analysis. In the categorization task, subjects had to decide whether stimuli represented natural objects or artefacts, a task placing higher demands on shape discrimination. The subjects also performed an object decision task, deciding whether pictures represented real objects or nonobjects. The nonobjects were chimeric line-drawings of closed figures constructed by exchanging single parts belonging to objects from the same category, which makes elaborate shape processing essential to perform this task (Gerlach et al., 2004, 2005). In addition to PET-measurements, we also recorded reaction times (RTs) in all tasks. Our general expectation was that demands on shape processing would influence rCBF in the VWFA. Because shape processing is required to different degrees in colour decision (where processing of shape is not necessary to perform the task) and categorization (where shape must be processed), we expected differences in rCBF in the VWFA, as well as in RTs between the tasks with pictorial stimuli. In the word-tasks, we expected the difference in rCBF between colour decision and categorization to be smaller, as reading–and thus the processing of word shape–is assumed to occur automatically, regardless of task demands (cf. the Stroop-effect, Stroop, 1935). Areas associated with conscious reading have also been found activated when words are presented subliminally (Dehaene et al., 2001), indicating elaborated processing even when subjects are unaware of seeing a word.

Thus we hypothesized that: (i) rCBF in the VWFA would be higher for words than for pictures during colour decision; (ii) this difference would diminish, as demands on shape processing increased, during categorization; and (iii) rCBF in the VWFA would be highest during object decision. We also expected reaction times to reflect task difficulty, so that RTs would be slowest in the object decision task, intermediate during categorization and comparatively fast in the colour decision task.

Method

Subjects

12 right-handed healthy volunteers (6 female, mean age 24 (range: 22–28)) participated in this experiment. Informed written consent was obtained according to the Declaration of Helsinki II, and the study was approved by the local ethics committee of Copenhagen (J.nr. (KF) 11-100/00).

Cognitive tasks

The subjects completed three tasks: (i) colour decision (deciding whether stimuli were coloured white or yellow), (ii) categorization (deciding whether stimuli represented natural objects or artefacts), and (iii) object decision (deciding whether stimuli represented real objects or non-objects). In the colour decision and categorization tasks, stimuli were either words or pictures (in separate conditions). The object decision task was with pictures only. All tasks came in two versions: one with a predominance of artefacts presented in the critical scan window and the other with a predominance of natural objects (cf. the section on design). Altogether, subjects were presented with a total of 10 conditions (see Table 1). In addition, the subjects also performed four other tasks which will not be considered here as they have no bearing on the present experiment (results have been presented in Gerlach et al., 2006). Because the purpose of this paper is to examine effects of stimulus type (words vs. pictures) and shape analysis in the VWFA, the behavioural and imaging analyses were collapsed over category (artefacts and natural objects).

The order of tasks and conditions was randomized across subjects with the only constraint that all subjects first performed the object decision task. The subjects were encouraged to respond as fast and as accurately as possible. Before the actual experiments started, the subjects performed a practice version of each task while in the scanner. Stimuli from these practice versions were not used in the actual experiments.

All responses were made on a serial response box placed in front of the subjects' right hand. Half of the subjects used their index finger to answer for yellow stimuli (colour decision), artefacts (categorization) and real objects (object decision) and

Table 1
Cognitive tasks

Task	Condition	Stimuli
Colour decision	Pictures	Natural objects
		Artefacts
	Words	Natural objects
		Artefacts
Categorization	Pictures	Natural objects
		Artefacts
	Words	Natural objects
		Artefacts
Object decision	Pictures	Natural objects
		Artefacts

their middle finger for the other responses. The other half of the subjects did the reverse.

Design

Seventy stimuli were presented in each condition. Half of these were drawn in yellow whereas the other half were drawn in white. All stimuli were presented on a black background on a PC-monitor hanging 60 cm in front of the subjects. The stimuli subtended between 3° and 5° of visual angle and were presented in the centre of gaze. Each stimulus was displayed for 180 ms, with an inter-stimulus-interval of 1320 ms, making each task last 1 min and 45 s. All tasks were initiated approximately 1 min and 15 s prior to isotope arrival to the brain and continued during the first 30 s of acquisition corresponding to the delivery of radio-tracer to the brain. From the point of task offset, the subjects viewed a blank screen for the next 60 s, yielding a total acqui-sition time of 90 s.

During both colour decision and categorization, 35 stimuli representing artefacts and 35 stimuli representing natural objects were presented. The presentation of stimuli in each condition was blocked in two: In one categorization task with pictures, the first block consisted of 19 natural objects and 32 artefacts, whereas the second block consisted of 16 natural objects and 3 artefacts. In the other categorization task with pictures the first block consisted of 19 artefacts and 32 natural objects whereas the second block consisted of 16 artefacts and 3 natural objects. The categorization task with words and the colour decision task with words and pictures were arranged in a similar manner. The order of the stimuli (yellow vs. white/artefact vs. natural object) was randomized within each block.

Each version of the object decision task consisted of line-drawings of 35 real objects and 35 nonobjects. They were blocked in the same way as the categorization and colour decision tasks so that the first block consisted of 19 real objects (either natural objects or artefacts) and 32 nonobjects, whereas the second block consisted of 16 real objects (either natural objects or artefacts) and 3 nonobjects. The order of the pictures (real vs. nonobject) was randomized within each block.

In all tasks the two blocks were presented sequentially but arranged so that the first block would be initiated approximately 45 s before injection and last until the bolus was estimated to reach the brain. The second block was displayed in the actual uptake phase of the tracer and ended before washout was likely to begin (the critical scan window).

Stimuli

The line drawings of real objects were selected from various sources but mainly from the standardised set of Snodgrass and Vanderwart (1980). The pictures used in the second block were all selected from the pool of Snodgrass and Vanderwart (1980). Six sets of objects (three sets of natural objects and three sets of artefacts), with 16 items in each set, were selected. The words used for the colour decision and categorization tasks corresponded to the names of the objects presented in the picture conditions. The nonobjects used in the object decision task were selected mainly from the set made by Lloyd Jones and Humphreys (1997). Because these nonobjects are composed of parts of objects from the same category, they can be considered either natural or artefactual. In the object decision task one particular category of nonobject was paired with its corresponding real object category. Six different sets of nonobjects were used (three sets of natural nonobjects and three sets of artefactual nonobjects).

Given that six different sets of concepts were used (three sets with concepts representing artefacts and three sets with concepts representing natural objects), these sets were rotated across subjects so that any given concept would never be presented twice for the same subject. In this way repetition effects were avoided.

PET scanning

PET scans were obtained with an 18-ring GE-Advance scanner (General Electric Medical Systems, Milwaukee, WI, USA) operating in 3D acquisition mode, producing 35 image slices with an interslice distance of 4.25 mm (for technical specifications see DeGrado et al., 1994). Each subject received 14 intravenous bolus injections of 200 MBq of $H_2^{15}O$ with an interscan interval of 8–10 min. The isotope was administered in an antecubital intravenous catheter over 20 s by an automatic injection device followed by 10 ml of physiological saline for flushing. Head movements were limited by head-holders constructed by thermally moulded foam. Before the activation sessions a 10 min. transmission scan was performed for attenuation correction. Images were reconstructed using a 4.0 mm Hanning filter transaxially and an 8.5 mm Ramp filter axially. The resulting distribution images of time integrated counts were used as indirect measurements of the regional neural activity.

Image analysis

For all subjects the complete brain volume was sampled. Image analysis was performed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Cognitive Neurology, London, UK). All intra-subject images were aligned on a voxel-by-voxel basis using a 3-D automated six parameters rigid body transformation. The average PET scans were subsequently transformed into the standard stereotactic atlas of Talairach and Tournoux (1988) using the PET template defined by the Montreal Neurological Institute (Friston et al., 1995a). The stereotactically normalized images consisted of 68 planes of $2 \times 2 \times 2$ mm voxels. Before statistical analysis, images were filtered with a 12 mm isotropic Gaussian filter to increase the signal-to-noise ratio and to accommodate residual variability in morphological and topogra-phical anatomy that was not accounted for by the stereotactic normalization process (Friston, 1994). Differences in global activity were removed by proportional normalization of global brain counts to a value of 50.

Tests of the null hypothesis, which rejects regionally specific condition activation effects, were performed comparing conditions on a voxel-by-voxel basis. The resulting set of voxel values constituted a statistical parametric map of the t-statistic, SPM{t}. A transformation of values from the SPM{t} into the unit Gaussian distribution using a probability integral transform allowed changes to be reported in Z-scores (SPM{Z}). The voxel significance threshold was estimated using the theory of Gaussian fields (Friston et al., 1995b; Worsley et al., 1996).

Given the objective of the present study, we restricted our analysis to the putative VWFA as defined by Cohen and Dehaene (2004). According to these authors the VWFA is found at the same location in Talairach space (approximately: $x, y, z = -43, -54, -12$) with a standard deviation of about 5 mm. Hence, we used a small

volume correction centred on this coordinate with a radius of 5 mm. We used this ROI in the following way: First, we performed the contrasts of interest using whole brain analysis. The threshold for these contrasts was set at $P_{uncorrected} < 0.05$ ($t > 1.66$). We then adjusted the results from these contrasts based on the small volume correction with a threshold of $P_{corrected} < 0.05$.

Results

Behavioural data

All the analyses presented are based on the mean correct RTs to the 16 stimuli presented in the critical scan window.

The data from the colour decision and categorization tasks were subjected to a two-by-two factorial analysis. The factors were Task Type with two levels (colour decision vs. categorization) and Stimulus Type also with two levels (words vs. pictures). This analysis revealed a significant main effect of Task Type [$F (1, 11) = 54.7$, $P < 0.001$], with faster responses to objects presented in the colour decision task, a marginally significant main effect of Stimulus Type [$F (1, 11) = 4.4$, $P < 0.06$], with faster responses to pictures than words, and a significant interaction between Task Type and Stimulus Type [$F (1, 11) = 27.5$, $P < 0.001$]. Post-hoc analyses (Tukey HSD tests) of the main effects revealed a significant difference ($P < 0.05$) in responses to both pictures and words in colour decision vs. categorization, and in responses to words and pictures during categorization. There was no significant difference between responses to words and pictures during colour decision. Together, these results indicate that the effect of Stimulus Type on RT is only evident during categorization, where pictures are processed faster than words. Analyses of errors revealed a significant difference in error rate between colour decision and categorization with words, with more errors in the categorization task (Wilcoxon Signed Ranks test: $P < 0.05$), whereas there was no significant difference for pictures (Wilcoxon Signed Ranks test: $P = 0.14$). Also, significantly more errors were made with words than pictures during categorization (Wilcoxon Signed Ranks test: $P < 0.05$), whereas there was no significant difference in error rate for words and pictures during colour decision (Wilcoxon Signed Ranks test: $P = 0.13$).

To examine whether object decisions were more difficult to perform than categorizations, RTs for object decisions were compared with RTs for the categorization of pictures. This analysis revealed a trend with RTs tending to be slower during object decision compared with categorization (Paired sample t-test: $P = 0.094$). Analyses of errors revealed that more errors were made during object decision compared with categorization (Wilcoxon Signed Ranks test: $P < 0.05$). The mean correct RTs and the mean
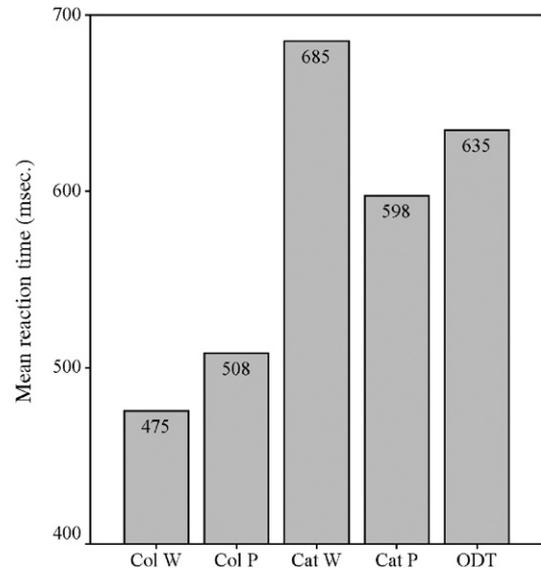

Fig. 1. Plot of the mean correct reaction time for each of the five tasks. Col W = Colour decision task with words, Col P = Colour decision task with pictures, Cat W = Categorization task with words, Cat P = Categorization task with pictures, ODT = Object decision task.

and median error rates, collapsed over category, for all tasks are given in Table 2 (see also Fig. 1).

Imaging data

The data from the colour decision and the categorization tasks were subjected to a two-by-two factorial analysis. The factors were Task Type with two levels (colour decision vs. categorization) and Stimulus Type with two levels (words vs. pictures). This analysis revealed a significant main effect of Task Type associated with increased rCBF during categorization compared to colour decision [$(x, y, z) = (-42, -56, -16)$; $Z = 5.18$; $P_{corrected} = 0.001$] and a significant main effect of Stimulus Type associated with increased rCBF during processing of words compared with pictures [$(x, y, z) = (-48, -56, -12)$; $Z = 2.86$; $P_{corrected} = 0.022$]. No other effects were significant.

Despite that no interaction was found between Task Type and Stimulus Type, inspection of blood flow values for the coordinate $(x, y, z) = (-48, -56, -12)$ (Fig. 2A) indicates that the effect of Stimulus Type is more pronounced during colour decision than during categorization. To address this difference more formally, we broke the analysis into two separate analyses; one for the colour decision task and one for the categorization task. This procedure confirmed that the effect of Stimulus Type was stronger during colour decisions than during categorizations, as the effect of Stimulus Type was still significant in the analysis restricted to colour decision [$(x, y, z) = (-48, -56, -10)$; $Z = 3.03$; $P_{corrected} = 0.017$] but not in the analysis restricted to categorization.

Discussion

The imaging data revealed a main effect of task type in the left fusiform gyrus, with higher rCBF during categorization compared to colour decision. A similar effect was found in the behavioural

Table 2
Mean reaction times (S.D.s) and mean (median) error rates for the 16 objects presented in the second block of the experimental tasks

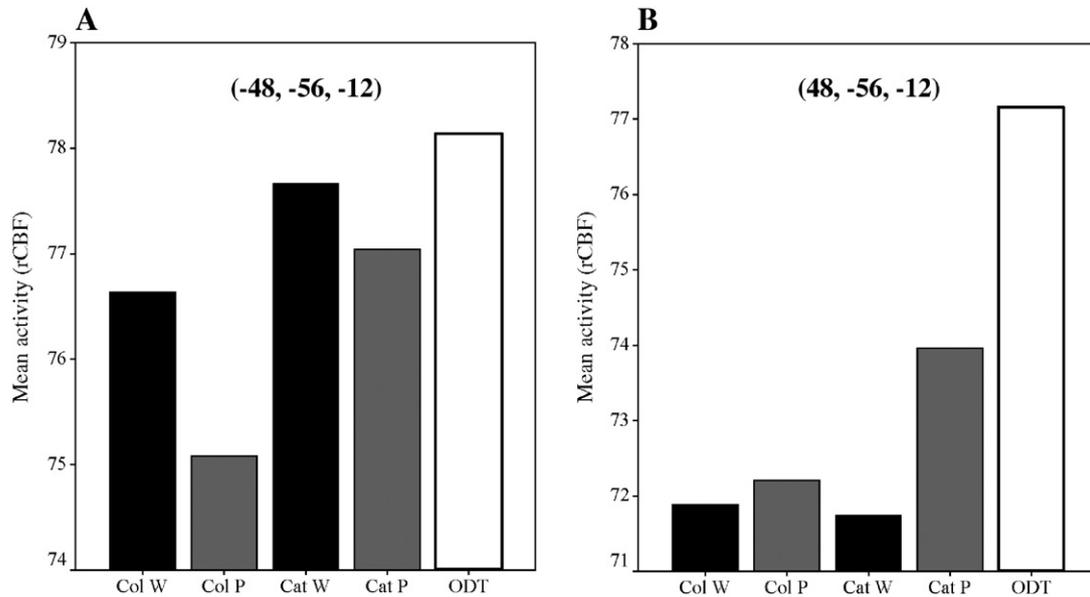|	Mean correct RT (S.D.)	Mean error rate (Median)
Colour decision words	475 (81)	0.6 (0)
Colour decision pictures	508 (84)	0.2 (0)
Categorization words	685 (77)	1.1 (1)
Categorization pictures	598 (86)	0.5 (0)
Object decision	635 (89)	1.3 (1)

Fig. 2. Plot of the mean activity for each of the five tasks in (A) the *left* VWFA and (B) the *right* homologue of the VWFA. Col W = Colour decision task with words, Col P = Colour decision task with pictures, Cat W = Categorization task with words, Cat P = Categorization task with pictures, ODT = Object decision task. From these plots it is clear that the effect of stimulus type (words vs. pictures) in the left VWFA diminishes as demands on shape analysis increase. It is also clear that there is no increased response to words in the right VWFA where rCBF is generally higher for pictures. The S.D.s for each condition in the left VWFA were as follows: Col W = 3.6, Col P = 3.7, Cat W = 3.6, Cat P = 3.8, and ODT = 4.0. For the r-VWFA they were: Col W = 5.7, Col P = 5.4, Cat W = 4.9, Cat P = 5.7, and ODT = 6.2. The plot for the r-VWFA was obtained by extracting blood flow values for the coordinate $(x, y, z) = (48, -56, -12)$ from the *F*-contrast.

data where RTs were longer for categorizations than for colour decisions. These findings support our initial assumption that categorizations require more detailed processing than colour decisions.

The imaging data also revealed a main effect of stimulus type in the left mid fusiform gyrus, with higher rCBF during word compared with picture processing in the context of both colour decision and categorization. When compared individually for colour decision and categorization, the difference between words and pictures was only significant during colour decision, although a similar trend was present during categorization also. Given that the left fusiform gyrus is likely to support shape processing, as we will argue below, and given that a greater degree of shape analysis is demanded for categorization compared to colour decision, this suggests that the effect of stimulus type is reduced as demands on shape processing increase.[1] This interpretation is supported by the observation that rCBF is higher in this region during object decision compared with both categorization of words ($Z = 2.55$; $P_{corrected} = 0.046$) and pictures ($Z = 3.35$; $P_{corrected} = 0.006$). This finding is in line with previous studies which have shown that object decision tasks require more fine-grained shape analysis than superordinate categorization tasks with pictures, causing greater activation in posterior and ventral

parts of the brain including the fusiform gyri (Gerlach et al., 2000, 2004). This explanation is also consistent with the observation that rCBF in the left mid fusiform gyrus is lower during categorization of pictures than during categorization of words (see Fig. 2A). To categorize a picture, only a few shape characteristics may need to be processed, whereas for words the majority of letters must be processed before the word can be categorized. This interpretation may even account for the effect of stimulus type found in the behavioural data. Here the effect of stimulus type was only significant during categorization where pictures were categorized faster than words. A picture superiority effect during categorization has been reported several times (e.g., Pellegrino et al., 1977; Potter and Faulconer, 1975; Snodgrass and McCullough, 1986) and may reflect that pictures have a privileged access to semantics (Caramazza et al., 1990): Pictures have structural features that correspond to semantic features, whereas this is not the case for words which, in this respect, have completely arbitrary shapes.

That rCBF increases in the putative VWFA as a function of the degree of shape analysis required, especially for pictures, and that the effect of stimulus type is diminished as the demand on shape analysis increases, clearly suggest a role for this area in operations besides word processing. It is also clear, however, that there is something special about words in this area: (i) when the demand for shape analysis is low, rCBF *is* higher for words than for pictures, (ii) when the demand for shape analysis increase during categorization a tendency in the same direction is still present, and (iii) there is no similar trend in the homologue area of the right hemisphere (compare Figs. 2A, B). Accordingly, there appears to be two effects present in the activation pattern in the VWFA, namely effect of stimulus type and of demand on shape analysis.

---

[1] These findings are unlikely to reflect the specific kernel size used in the preprocessing of the imaging data (12 mm) as we found similar results with kernel sizes of 10 and 8 mm (with a kernel size of 6 the main effect of stimulus type was still significant but none of the simple main effects were, preventing any meaningful interpretation of their individual magnitude). Hence, it is unlikely that an increase in rCBF in this region specific to words just happened to be overshadowed (due to smoothing) by an even larger increase in rCBF to pictures in surrounding regions.

Same same but different

We have suggested that rCBF in the putative VWFA increases as a function of the degree of shape analysis required, but what specific role might the mid fusiform serve in shape processing?

There is reason to believe that activation in the left mid fusiform gyrus reflects structural rather than semantic processing. It has been shown that activation in this region is not modulated by semantic parameters such as familiarity (Whatmough et al., 2002), and that areas associated with semantic processing during super-ordinate categorization are located much more anteriorly in the temporal cortex (Gerlach et al., 2000). Although it has been suggested that the VWFA might serve as a link between structural descriptions of words and representations in long-term memory (Kronbichler et al., 2004, but see Devlin et al., 2006 for a discussion), there is little evidence pointing at this area serving such a role in object recognition. Areas associated with stored visual object representations are probably located more anteriorly and ventrally in the fusiform gyri (Bar et al., 2001; Gerlach et al., 2002, 2006; Joseph and Gathers, 2003). Instead, we suggest that activation in this area is likely to reflect shape configuration, that is, the integration of shape elements into more elaborate shape descriptions corresponding to whole objects or large object parts (Gerlach et al., 2002, 2006). This can account for the present observation that rCBF in this region is lowest during colour decision, intermediate during categorization and highest during object decision. Object decision requires object individuation–and therefore formation of detailed shape descriptions–whereas categorization does not to the same degree. During categorization, it may be sufficient to recognize single features to judge something as belonging to a certain category. The same approach will not suffice to judge something as a real object in the object decision task because the non-objects are created by real objects from the same category, and thus contain the same features. With respect to the colour decisions, these can in principle be performed without need for shape configuration at all.

What must be explained then, is how shape configuration can give rise to an effect of stimulus type. If we consider pictures, these can be identified at different levels going from a subordinate level (e.g., bulldog), over a base level (dog), to a superordinate level (animal). The more specific the level of identification required, the more detailed shape processing must be (Gauthier et al., 2000; Kosslyn et al., 1995; Rogers et al., 2005). The same is not true of words. Furthermore, word processing is such an automated skill that people cannot help reading a word when it is within their field of attention even if they are told not to, as witnessed by the Stroop effect (Stroop, 1935). Also, when subjects perform a non-linguistic task with written word stimuli, classical language areas–including the VWFA–are activated (Price et al., 1996). Accordingly, while both word and picture processing require shape configuration (for word identification, letters and their constituent parts need to be combined, and for objects, details must be integrated), the two stimulus types may differ in the degree of shape configuration they are subjected to. For pictures, the degree of shape configuration performed may vary as a function of the task (e.g., subordinate, base, or superordinate level identification) whereas for words the amount of shape configuration performed is likely to be rather constant as long as the word is read. Although automatic processing has been associated with stimulus specificity and anatomical localization (Fodor, 1983), we are not proposing that the more automatic processing of words in this area reflects

modularity. On the contrary, we are arguing that the process of shape configuration is common to words and pictures, but differently affected by task demands.

Based on these premises it is possible to account for the activation pattern observed in the putative VWFA. During the colour decision task the amount of shape configuration undertaken is flexible for pictures but less so for words, and because shape configuration is in principle not demanded in this task, rCBF is higher during word processing than during picture processing. In the categorization task the need for shape configuration increases for pictures, but not as much for words, causing the effect of stimulus type to diminish. Lastly, during object decision, the need for shape configuration is so high for pictures that it will abolish any shape configuration effect on rCBF for words.[2]

One thing left to be explained is why we do not see the same pattern in the right hemisphere homologue of the putative VWFA (r-VWFA). It has been suggested that the r-VWFA might assume the same role as the VWFA, but that this is not necessary for normal reading (Cohen et al., 2003). Activation of the r-VWFA by alphabetic stimuli is only consistently found when comparing words to fixation (Cohen et al., 2002). Compared to other visual stimuli like checkerboards (Cohen et al., 2003), words do not generate more activity in the r-VWFA. Furthermore, lesions in this area do not give rise to alexia. Any specific explanation for this left-lateralization remains speculative, but in general terms, and as argued by Cohen et al. (2000, 2003), it is likely to reflect that processing in the VWFA is at a level of specialization where language dominance is important. Another possibility is that the right and left hemispheres play different parts in shape configuration. Processing of local shape characteristics is primarily supported by the left hemisphere, while the right hemisphere is biased towards processing of global shape characteristics (Fink et al., 1996, 1997; Lamb et al., 1989, 1990). On this basis, one could speculate that the global shapes of words may not provide many clues to word identity, as reading is based on identification of differences primarily at a local (letter or feature) level (see Pelli et al., 2003 for evidence supporting this notion). The same is not to true of objects because global shape characteristics are, to some extent, diagnostic of object identity (Gerlach et al., 2006).

Shape configuration in r-e-a-d-i-n-g

If the left mid fusiform gyrus is important for shape configuration, patients with lesions including this area should have difficulties with fast word recognition. This is in accordance with data from patient studies where lesions in this region are associated with letter-by-letter reading (Cohen et al., 2003; Gaillard et al., 2006; Leff et al., 2006). Such a reading strategy would be natural to employ if coherent word representations cannot be derived efficiently, which they cannot if shape configuration fails. We are not arguing that all patients with letter-by-letter reading should have lesions in this area, or that letter-by-letter reading is

---

[2] We acknowledge that we cannot say for certain that the word effect will be completely abolished at increasing levels of shape analysis. This would require that the word effect failed to show up in a word processing task that was as demanding as the most demanding object processing task used here (object decision). While we suspect that the effect would be eliminated, as the effect of stimulus type was already reduced in the categorization task, we are not able to test this prediction here.

always a consequence of impaired shape configuration.[3] What we are suggesting is that patients with lesions in the VWFA should invariably have reading problems affecting the integration of shapes (letters or letter features) into words, and that these patients will be likely to resort to the compensating strategy of letter-by-letter reading.

Our suggestion implies that lesions affecting this area should not only impair word processing but also visual object processing, as shape configuration is involved in both. Evidence compatible with this suggestion comes from studies showing impaired visual object processing in patients with letter-by-letter reading following ventral occipito-temporal lesions (Behrmann et al., 1998a,b; Farah and Wallace, 1991; Sekuler and Behrmann, 1996).

This suggestion is not incompatible with reports of accurate object naming in patients with lesions in the putative VWFA (e.g., Gaillard et al., 2006), as the visual object processing abilities of such patients have typically been assessed in picture naming tasks with no reaction time measurement. Patients can exhibit quite accurate object naming performance despite marked impairments in visual object processing (Davidoff and Warrington, 1999) including impaired shape configuration (Gerlach et al., 2005). Hence, tasks have to be quite challenging for object processing deficits to be revealed, and studies employing more fine grained measures do report impaired visual object processing in patients with letter-by-letter reading. As an example, the patients studied by Behrmann et al. (1998a) only exhibited a clear object processing deficit with pictures of high visual complexity. This, however, does not explain why tasks need to be more demanding for objects than for words to reveal the impairment in object processing. On the account suggested here, the explanation could be that shape configuration is not necessary for e.g. accurate picture naming at the basic level (although it may be for efficient object naming) because accurate object naming can be based on only a few shape characteristics. Hence, only in tasks that place high demands on shape configuration, or in tasks where RT is measured, will an impairment become apparent (see Gerlach et al., 2005). This is not so for words which, according to the present account, undergo similar degrees of shape configuration regardless of task demands.

The suggestion that the putative VWFA contributes to both object processing and word processing is also compatible with imaging studies, besides the present, that have found activation in this area during picture processing (e.g., Moore and Price, 1999). Using stringent criteria for selectivity, Joseph et al. (2006) failed to find any letter specific activation in the fusiform gyrus, when comparing letters and objects in two different tasks (matching and naming). They concluded that "this region subserves a cognitive process shared by both letters and objects" (p. 815) and we suggest that this shared cognitive process is shape configuration.

In defence of labelling the left mid fusiform gyrus as the Visual Word Form Area, Cohen and Dehaene (2004) have argued that the overlap observed across studies examining picture and word processing may be a consequence of the variability in the location of the VWFA across subjects (in different studies) as well as the degree of smoothing applied in the imaging analysis. In the present study, which is based on analysis in a region centred on the co-ordinates for the VWFA as given by Cohen and Dehaene (2004), we do find greater activation for words than for pictures—but only in the colour decision condition. When demands on shape processing increase, this effect is no longer significant and picture processing during object decision causes greater increase in rCBF in the VWFA than word processing. Given that this pattern of activation is found with the same subjects and across different levels of smoothing (see footnote 1), the overlap noticed in previous studies may signify a genuine involvement of this area in both word and picture processing.

Conclusion

We examined rCBF in the left mid fusiform gyrus–in the region labelled the Visual Word Form Area–during processing of words and pictures and in tasks that varied in demands placed on shape processing. We find that rCBF in this area is modulated by both stimulus type (words vs. pictures) and task demands. rCBF is significantly greater during word than during picture processing, but only in tasks that do not require elaborate processing of shape. When the demand on shape processing increase this difference is reduced, and the rCBF during picture processing may even exceed that observed during word processing. These findings suggest that activation in this region is not specific to words, but rather reflects an operation common to word and picture processing that may be differentially affected by task demands. We suggest that activation in this area could reflect shape configuration–the integration of shape elements into more elaborate shape descriptions corresponding to whole objects or words–which may be performed to different degrees for pictures and words depending on task requirements. Shape configuration of words might occur to a similar degree regardless of the task at hand, while for pictures this process may be more dependent on the level of analysis required in the specific task. This can also explain why the VWFA is quite consistently found activated in studies comparing words to simple visual stimuli, whereas the findings are more diverging when words are compared to pictures or faces.

While the observed pattern of rCBF in the left middle fusiform gyrus can be explained by the construct 'shape configuration', this aspect of visual object processing has not been examined to any large degree (Behrmann and Kimchi, 2003). Further studies are needed to explore this aspect of visual analysis, and how it may contribute to the processing of words and pictures. The cerebral substrate of shape configuration of different stimuli also deserves further investigation.

Acknowledgments

References

Bar, M., Tootell, R.B., Schacter, D.L., Greve, D.N., Fischl, B., et al., 2001. Cortical mechanisms specific to explicit visual object recognition. Neuron 29, 529–535.

Behrmann, M., Kimchi, R., 2003. Visual perceptual organization: lessons from lesions. In: Kimchi, R., Behrmann, M. (Eds.), Perceptual

---

[3] Although the terms pure alexia and letter-by-letter reading have to a large extent been used interchangeably in the literature, we acknowledge that letter-by-letter reading is a compensating strategy available to patients with different kinds of reading deficits with different aetiologies. In the current discussion the terms are used interchangeably, and letter-by-letter reading thus refers to the use of this strategy by patients with pure alexia.

Organization in Vision: Behavioral and Neural Perspectives. Lawrence Erlbaum Associates Publishers, London, pp. 337–375.

Behrmann, M., Nelson, J., Sekuler, E.B., 1998a. Visual complexity in letter-by-letter reading: "pure" alexia is not pure. Neuropsychologia 36, 1115–1132.

Behrmann, M., Plaut, D.C., Nelson, J., 1998b. A literature review and new data supporting an interactive account of letter-by-letter reading. Cogn. Neuropsychol. 15, 7–51.

Caramazza, A., Hillis, A.E., Rapp, B.C., Romani, C., 1990. The multiple semantics hypothesis: multiple confusions? Cogn. Neuropsychol. 7, 161–189.

Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. NeuroImage 22, 466–476.

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

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., et al., 2002. Language-specific tuning of visual cortex? Functional properties of the visual word form area. Brain 125, 1054–1069.

Cohen, L., Martinaud, O., Lemer, C., Lehericy, S., Samson, Y., et al., 2003. Visual word recognition in the left and right hemispheres: anatomical and functional correlates of peripheral alexias. Cereb. Cortex 13, 1313–1333.

Davidoff, J., Warrington, E.K., 1999. The bare bones of object recognition: implications from a case of object recognition impairment. Neuropsychologia 37, 279–292.

DeGrado, T.R., Turkington, T.G., Williams, J.J., Stearns, C.W., Hoffman, J.M., et al., 1994. Performance characteristics of a whole-body PET scanner. J. Nucl. Med. 35, 1398–1406.

Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.F., Poline, J.B., Rivière, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. 4, 752–758.

Devlin, J.T., Jarnison, H.L., Gonnerman, L.M., Matthews, P.M., 2006. The role of the posterior fusiform gyrus in reading. J. Cogn. Neurosci. 18, 911–922.

Farah, M.J., Wallace, M.A., 1991. Pure alexia as a visual impairment: a reconsideration. Cogn. Neuropsychol. 8, 313–334.

Fink, G.R., Halligan, P.W., Marshall, J.C., Frith, C.D., Frackowiak, R.S., Dolan, R.J., 1996. Where in the brain does visual attention select the forest and the trees? Nature 382, 626–628.

Fink, G.R., Halligan, P.W., Marshall, J.C., Frith, C.D., Frackowiak, R.S., Dolan, R.J., 1997. Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. Brain 120, 1779–1791.

Fodor, J.A., 1983. The Modularity of Mind. MIT press, Cambridge, MA.

Friston, K.J., 1994. Statistical parametric mapping. In: Thatcher, R.W., Hallett, M., Zeffiro, T., John, E.R., Huerta, M. (Eds.), Functional Neuroimaging. Academic Press, San Diego, pp. 79–93.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., et al., 1995a. Spatial registration and normalization of images. Hum. Brain Mapp. 3, 165–189.

Friston, K.J., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995b. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., et al., 2006. Direct intracranial fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. Neuron 50, 191–204.

Gauthier, I., Tarr, M.J., Moylan, J., Anderson, A.W., Skudlarski, P., Gore, J.C., 2000. Does visual subordinate-level categorisation engage the functionally defined fusiform face area? Cogn. Neuropsychol. 17, 143–163.

Gerlach, C., Law, I., Gade, A., Paulson, O.B., 2000. Categorization and category effects in normal object recognition: a PET study. Neuropsychologia 38, 1693–1703.

Gerlach, C., Aaside, C.T., Humphreys, G.W., Gade, A., Paulson, O.B., et al., 2002. Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. Neuropsychologia 40, 1254–1267.

Gerlach, C., Law, I., Paulson, O.B., 2004. Structural similarity and category-specificity: a refined account. Neuropsychologia 42, 1543–1553.

Gerlach, C., Marstrand, L., Habekost, T., Gade, A., 2005. A case of impaired shape integration: Implications for models of visual object processing. Vis. Cogn. 12, 1409–1443.

Gerlach, C., Law, I., Paulson, O.B., 2006. Shape configuration and category-specificity. Neuropsychologia 44, 1247–1260.

Hillis, A.E., Newhart, M., Heidler, J., Barker, P., Herskovits, E., et al., 2005. The roles of the "visual word form area" in reading. NeuroImage 24, 548–559.

James, K.H., James, T.W., Jobard, G., Wong, A.C.-N., Gauthier, I., 2005. Letter processing in the visual system: different activation patterns for single letters and letter strings. Cogn. Affect. Behav. Neurosci. 5, 452–466.

Joseph, J.E., Gathers, A.D., 2003. Effects of structural similarity on neural substrates for object recognition. Cogn. Affect. Behav. Neurosci. 3, 1–16.

Joseph, J.E., Gathers, A.D., Piper, G.A., 2003. Shared and dissociated cortical regions for object and letter processing. Brain Res. Cogn. Brain Res. 17, 56–67.

Joseph, J.E., Cerullo, M.A., Farley, A.B., Steinmetz, N.A., Mier, C.R., 2006. fMRI correlates of cortical specialization and generalization for letter processing. NeuroImage 32, 806–820.

Kosslyn, S.M., Alpert, N.M., Thompson, W.L., 1995. Identifying objects at different levels of hierarchy: a positron emission tomography study. Hum. Brain Mapp. 3, 107–132.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., et al., 2004. The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. NeuroImage 21, 946–953.

Lamb, M.R., Robertson, L.C., Knight, R.T., 1989. Attention and interference in the processing of global and local information: effects of unilateral temporal–parietal junction lesions. Neuropsychologia 27, 471–483.

Lamb, M.R., Robertson, L.C., Knight, R.T., 1990. Component mechanisms underlying the processing of hierarchically organized patterns: inferences from patients with unilateral cortical lesions. J. Exp. Psychol. Mem. Cogn. 16, 471–483.

Leff, A.P., Spitsyna, G., Plant, G.T., Wise, R.J.S., 2006. Structural anatomy of pure and hemianopic alexia. J. Neurol. Neurosurg. Psychiatry. 77, 1004–1007.

Lloyd Jones, T.J., Humphreys, G.W., 1997. Perceptual differentiation as a source of category effects in object processing: evidence from naming and object decision. Mem. Cogn. 25, 18–35.

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.

Moore, C.J., Price, C.J., 1999. Three distinct ventral occipitotemporal regions for reading and object naming. NeuroImage 10, 181–192.

Pellegrino, J.W., Rosinski, R.R., Chiesi, H.L., Siegel, A., 1977. Picture–word differences in decision latency: an analysis of single and dual memory models. Mem. Cogn. 5, 383–396.

Pelli, D.G., Farell, B., Moore, D.C., 2003. The remarkable inefficiency of word recognition. Nature 423, 752–756.

Potter, M.C., Faulconer, B.A., 1975. Time to understand pictures and words. Nature 253, 437–438.

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., 2004. The pro and cons of labelling a left occipitotemporal region: "the visual word form area". NeuroImage 22, 477–479.

Price, C.J., Wise, R.J.S., Frackowiak, R.S.J., 1996. Demonstrating the implicit processing of visually presented words and pseudowords. Cereb. Cortex 6, 62–70.

Price, C.J., Gorno-Tempini, M.L., Graham, K.S., Biggio, N., Mechelli, A., et al., 2003. Normal and pathological reading: converging data from lesion and imaging studies. NeuroImage 20 (Suppl. 1), 30–41.

Puce, A., Allison, T., Asgari, M., Gore, J.C., McCarthy, G., 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J. Neurosci. 16, 5205–5215.

Rogers, T.T., Hocking, J., Mechelli, A., Patterson, K., Price, C., 2005. Fusiform activation to animals is driven by the process, not the stimulus. J. Cogn. Neurosci. 17, 434–445.

Sekuler, E.B., Behrmann, M., 1996. Perceptual cues in pure alexia. Cogn. Neuropsychol. 13, 941–974.

Snodgrass, J.G., McCullough, B., 1986. The role of visual similarity in picture categorization. J. Exper. Psychol., Learn., Mem., Cogn. 12, 147–154.

Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. Mem. 6, 174–215.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging. Thieme, Stuttgart.

Vigneau, M., Jobard, G., Mazoyer, B., Tzourio-Mazoyer, N., 2005. Word and non-word reading: what role for the visual word form area? NeuroImage 27, 694–705.

Whatmough, C., Chertkow, H., Murtha, S., Hanratty, K., 2002. Dissociable brain regions process object meaning and object structure during picture naming. Neuropsychologia 40, 174–186.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., et al., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.