Vision Research xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

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

The role of invariant line junctions in object and visual word recognition

Marcin Szwed [a,c,f,*], Laurent Cohen [b,c,e,f], Emilie Qiao [b], Stanislas Dehaene [a,c,d,g]

[a] INSERM, Cognitive Neuro-imaging Unit, IFR49 Gif sur Yvette, France
[b] Université Pierre et Marie Curie-Paris 6, Faculté de Médecine Pitié-Salpêtrière, IFR70 Paris, France
[c] CEA, NeuroSpin center, IFR49 Gif sur Yvette, France
[d] Collège de France, Paris, France
[e] AP-HP, Groupe hospitalier Pitié-Salpêtrière, Department of Neurology, Paris, France
[f] INSERM, UMR975, ICM Research Center, Paris, France
[g] Université Paris-Sud, IFR49, F-91191 Gif/Yvette, France

ARTICLE INFO

Article history:
Received 17 September 2008
Received in revised form 13 January 2009
Available online xxxx

Keywords:
Reading
Invariance
Visual recognition
Word recognition
Word form

ABSTRACT

Object recognition relies heavily on invariant visual features such as the manner in which lines meet at vertices to form viewpoint-invariant junctions (e.g. T, L). We wondered whether these features also underlie readers' competence for fast recognition of printed words. Since reading is far too recent to have exerted any evolutionary pressure on brain evolution, visual word recognition might be based on pre-existing mechanisms common to all visual object recognition. In a naming task, we presented partially deleted pictures of objects and printed words in which either the vertices or the line midsegments were preserved. Subjects showed an identical pattern of behavior with both objects and words: they made fewer errors and were faster to respond when vertices were preserved. Our results suggest that vertex invariants are used for object recognition and that this evolutionarily ancient mechanism is being co-opted for reading.

1. Introduction

Reading has been invented only ∼5400 years ago and there was no sufficient time or evolutionary pressure to develop a devoted brain system with a genetic basis. Consequently, reading must rely on pre-existing neural systems for vision and language, which may be partially co-opted or "recycled" for the specific problems posed by reading in a given script (Dehaene, 2005; Dehaene & Cohen, 2007; Kinzler & Spelke, 2007). In this paper, we ask to what extent and in which ways reading is based on recognition mechanisms initially evolved for visual object recognition.

Visual objects have certain invariant (or non-accidental) properties that are common to most viewpoints. These properties include the manner in which lines meet at vertices to form specific configurations such as T or L, also referred to as line junctions and line coterminations. For example, a table contains several T junctions where the legs join the table top, and these junctions are common to all but a few unusual viewpoints. It is well established that such invariant properties are particularly important for object recognition (Biederman, 1987, 1995, Gibson, 1979, Lowe, 1987; Pitts & McCulloch, 1947) and a number of studies have demonstrated the importance of line vertices for perception with modeling (Binford, 1981; Lowe, 1987), electrophysiological methods in primates (Brincat & Connor, 2004; Kayaert, Biederman, & Vogels, 2003), and behavioral methods in pigeons (Gibson, Lazareva, Gosselin, Schyns, & Wasserman, 2007; Lazareva, Wasserman, & Biederman, 2008) and humans (Biederman, 1987; Gibson et al., 2007; Lazareva et al., 2008).

Interestingly, while writing systems vary a great deal in character shape and complexity, one can also find a similarity of the elementary building blocks that make writing symbols. Letters and ideograms (such as Kanji words) are all composed of a small and relatively constant number of lines that meet at vertices (Changizi & Shimojo, 2005). Changizi, Zhang, Ye, and Shimojo (2006) also found that in all of the world's writing systems, vertex configurations such as T or L obey a universal distribution which is shared with that found in environmental images. The basic building blocks of writing systems may therefore correspond to the key features used for object recognition (Changizi et al., 2006). Thus, the shape of written words may have been culturally selected to match the pre-existing constraints of our visual system (Changizi et al., 2006; Dehaene, 2005; Dehaene & Cohen, 2007).

In a classical article on the role of invariant properties in human object recognition, Biederman (1987) started with line drawings of objects and removed an equal amount of contour either at their vertices or at their midsegments. He observed that subjects responded more slowly and made more errors for objects in which

* Corresponding author. Address: Inserm U562 Cognitive Neuroimaging Unit, CEA/SAC/DSV/DRM/NeuroSpin, Bât 145, Point Courrier 156, F-91191 GIF/YVETTE, FRANCE. Fax: +33 (0) 1 69 08 79 73.
E-mail address: mfszwed@gmail.com (M. Szwed).

0042-6989/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2009.01.003

vertices were removed. This evidence supported the hypothesis that viewpoint-invariant vertex configurations play a significant role in object recognition. In this study we ask whether the same invariant properties play an important role in the recognition of written words. We do it by presenting, in a single experiment, objects and words made either of vertices or of line midsegments.

2. Methods

2.1. Participants and experimental set-up

Thirty eight subjects (mean age 26 ± 5.7 years, mean ± SD, 23 women and 15 men) with normal or corrected-to-normal vision participated in the experiments. Experiments were undertaken with the understanding and written consent of each participant. Participants were seated in a dimly lit room. Words and objects were presented on a video monitor (800 × 600, 75 Hz) on a white background at a distance of 80 cm, and subjects were asked to name stimuli aloud.

2.2. Stimuli

Stimuli consisted of printed words and of line drawings of objects. They could be either intact or degraded by the removal of line fragments. Two modes of degradation were used, depending on the type of visual features which were preserved: in the 'vertex' variant', the line junctions were preserved (Fig. 1 A–B, left), while in the 'midsegment' variant they were suppressed (Fig. 1 A–B, right). In both cases, an equal amount of contour was preserved, either 35% or 55% of the original, resulting in a total of up to 5 versions of each stimulus: intact, vertex-35%, vertex-55%, midsegment-35%, and midsegment-55%.

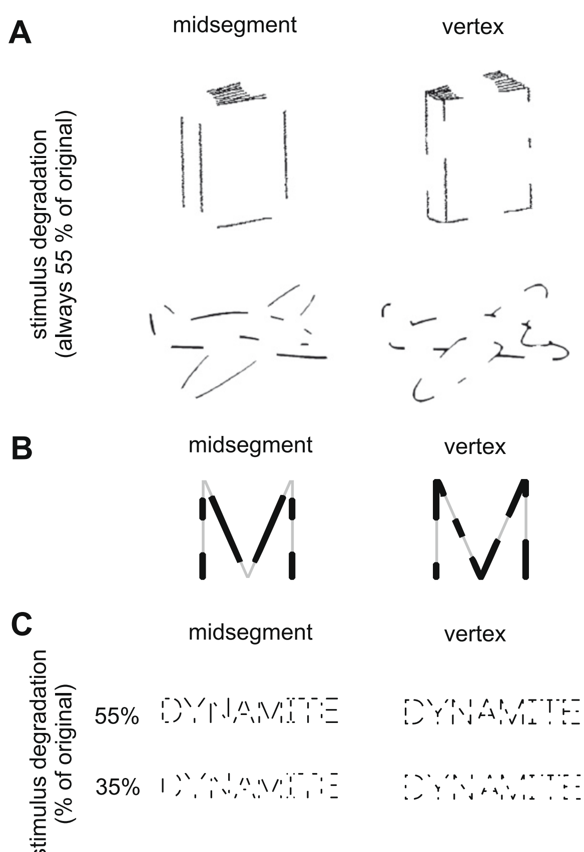

Fig. 1. Stimulus design. Subjects performed a naming task on partially degraded words and objects. Stimuli were degraded by partial deletion of some of their component lines, leaving intact either the vertex features or the midsegment features. (A) and (B), respectively, show sample objects and a sample letter, both with 55% of the original image preserved. The outline of the original letter is shown in thin light gray. Words were presented in fonts made of vertex and midsegment features, in "55% of original" (top) and "35% of original" (bottom) versions (C). Objects were always presented in "55% of original" version.

We selected vertices and line midsegments following the principles used by (Biederman, 1987) and (Changizi et al., 2006). We defined vertices as any junction of two or more lines. The transitions of straight lines into curves such as in the letter "J" were treated as vertices. We defined midsegments as line fragments at least 4 pixels away from any vertices. In the curvy parts of some letters, when distinct vertex and midsegment deletions could not be defined (e.g. anywhere in the letter "S"), identical deletions were made in the vertex and midsegments versions.

We attempted to keep the same number of deleted and preserved fragments across the 4 degraded versions of any given stimulus (Fig. 1). Deviations of 1 or 2 fragments more in either the vertex or midsegment version were allowed to preserve sufficiently the shape of stimuli. Since line terminations are very informative for letter recognition (Fiset et al., 2008) they were kept intact in virtually all letters (with the exception of 'A'). Objects subtended a visual angle of up to 3.9 × 4.6°. Words subtended a more elongated field of 0.8 × 5°. Fragment removal was implemented in Matlab (Mathworks, Natick, Massachusetts). Fonts were processed in Font Creator (High-Logic, Utrecht, Netherlands).

2.3. Objects

The object set included images from the Snodgrass and Vanderwart set (1980), images used by Lerner, Hendler, and Malach (2002), and a few additional images from children books. A total of 76 objects were used, 38 natural (e.g. animals, plants) and 38 artifacts (e.g. tools, clothes). When required, images were further simplified by removing textures and redundant details. Since line thickness varied substantially between images, we reduced it where necessary using filter commands in Photoshop (Adobe, San Jose, CA). We checked that the resulting images were still recognized at near 100% by running a pilot naming task.

2.4. Words

We used 6–8 letter French nouns with a frequency higher than one per million (www.lexique.org) (New, Pallier, Brysbaert, & Ferrand, 2004).

Letters such as C, O and S are made exclusively of curves that do not cross each other. Other letters, BDGJPQRU, are partially curvy. Our classification of features into vertices and midsegments, following those of Biederman (1987) and Changizi et al. (2006), remains agnostic about the role of such curvilinear features and manipulation of these curvy fragments is beyond the scope of this study. Therefore, we selected words made either exclusively or predominantly of 'non-curvy' letters (AEFHIKLMNTVWXYZ), allowing for either one 'fully curvy' or two 'partially curvy' letters (up to three partially curvy letters in case of eight-letter words).

We used an uppercase sans serif font with thin lines (Helvetica Ultra Light 42 points); a serif was added to the letter 'I'. We chose a line width and font size allowing us to equate satisfactorily luminance, line width, and line length across words and objects. Word and object sets were also matched in the number of vertices (5% difference in mean vertex count between words and objects).

2.5. Experimental design and data analysis

Each trial began with a 200 ms central fixation cross. It was then replaced by the target (either a word or an object), which remained

Please cite this article in press as: Szwed, M., et al. The role of invariant line junctions in object and visual word recognition. *Vision Research* (2009), doi:10.1016/j.visres.2009.01.003

on the screen for 200 ms (100 ms in some trials in Experiment 3). Participants were instructed to name the stimulus as quickly as possible while minimizing errors. No feedback was provided. The next trial started 1500 ms after the offset of the target.

Stimulus variants were counter-balanced between subjects; each subject saw any given stimulus in only one out of its five possible versions (e.g. a subject who saw the book in midsegment-55% variant, Fig. 1A left, would not see it in a vertex-55% variant, Fig. 1A right, nor in any of the other variants). Word and object trials were randomly intermixed. Responses were monitored online by the experimenter and recorded for offline analysis. Stimulation was implemented in E-prime 1.1 (PST, Pittsburgh, PA). Reaction times were acquired through a vocal key (PST Serial Response Box, PST, Pittsburgh, PA). Median RT were computed for each subject and each condition and entered in an ANOVA (or in equivalent paired t tests) with subjects as random factor. Error rates were analyzed using binary logistic regression with subjects as covariates. While in our case the distributions of error rates did not differ significantly from normal (Kolmogorov-Smirnov $p > 0.15$) we nonetheless applied binary logistic regression for the sake of statistical correctness (Baayen, 2004). All data were analyzed in E-prime (PST, Pittsburgh, PA), Microsoft Excel, Matlab, and Minitab (Minitab, State College, PA), except for the mixed-effect model which was implemented using the lmer function in the R package (www.r-project.org, Baayen, Davidson, & Bates, 2008).

2.6. Overview of experimental strategy

In *Experiment 1* ($n = 12$ subjects), we tested basic effects of the type of preserved feature (vertices or midsegments) on visual recognition. Subjects saw 180 words and 76 objects in either vertex-55% or midsegment-55% variants.

In *Experiment 2* ($n = 14$ subjects), we explored feature type effects in word perception in more depth. Subjects saw 420 words in intact, vertex-55%, vertex-35%, midsegment-55%, and midsegment-35% conditions (see Fig. 1). Object trials were the same as in Experiment 1.

In *Experiment 3* ($n = 12$ subjects), we probed the time course of the feature type effect by including an experimental condition with very short presentation time. Subjects saw 234 words in vertex-35% or midsegment-35% variants. In half of the trials, words were presented for 200 ms without masking (identical to Experiments 1 and 2). In the other half of the trials, words were presented for 100 ms and followed by a ####### mask that lasted 200 ms. No objects were shown.

In Section 3.1, the 'object' part is based on pooled results from Experiments 1 and 2. The 'word' part (subsequent sections) is based on results from Experiments 2 and 3; since the experimental conditions and words differed between the latter two, their results are always treated as separate data points.

3. Results

3.1. Effect of feature type on object recognition

Biederman (1987) found that line vertices were more important for object perception than line midsegments. Our first goal was to replicate this classical result using a set of simplified objects matched in luminance, line length and number of vertices to the word stimuli.

As Fig. 2A demonstrates, we found that subjects made significantly less naming errors for objects presented in the vertex variant (30% errors) than in the midsegment variant (43% errors) (binary logistic regression, $z = -6.25$, $p < .001$). RTs showed a parallel tendency, although the effect was not significant (vertex: 919 ms; midsegment; 928 ms). However, because there were sub-

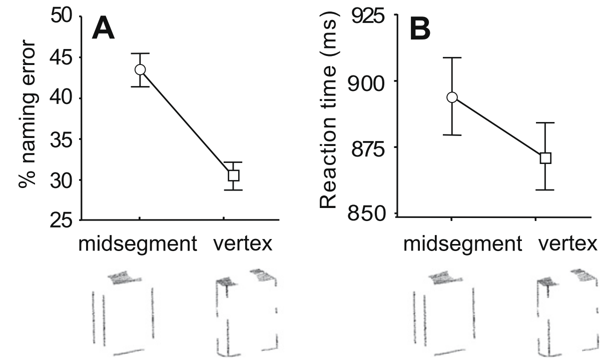

Fig. 2. Effect of feature type on object naming. (A) Error rates in a naming task were higher for objects presented in midsegment form (○) than in vertex (□) form. (B) Reaction times for a subset of 38 objects with low naming errors followed a similar but non-significant trend. Error bars denote S.E.M.

stantial differences in naming errors across individual objects (see section below) we reasoned that reaction time effects might be obscured by the fact that some objects were not recognized in the midsegment variant. Therefore, we repeated our analysis for the subset of 38 objects - half of the original set - which yielded fewer naming errors. The results are shown in Fig. 2B. We found that subjects were on average faster (28 ms) to respond to the vertex variant than to the midsegment variant (Fig. 2B). However, the trend was again not statistically significant ($t(25) = -.7$; $p = .12$).

We observed that not all objects suffered equally from being reduced to their vertices or midsegments. We therefore asked, in a subsequent analysis, what was the effect of feature type at the level of individual objects. For each picture, we computed error rates in the midsegment variant, in the vertex variant, and the difference of those two rates. Fig. 3A–C shows the corresponding distribution histograms. To illustrate this analysis we would like to consider the drawing of a book shown in Figs. 1 and 2. For the midsegment version, subjects made 77% errors, as opposed to only 8% errors for the vertex version, which makes a difference of 69%. This difference is considerably larger than the average error rate difference for the entire object set (13%). Inspection of the distribution of differences for all objects (Fig. 3C) shows that individual differences deviate substantially from the population mean (SD = 31%). In particular, this analysis revealed that, contrary to the average tendency, subjects performed worse with the vertex than with the midsegment version for 21 out of 76 (28%) objects. A typical example of such an object (a t-shirt) is shown in Fig. 3D.

In summary, we found that on average vertices are more important for object perception than line midsegments (Fig. 2). However, there are pronounced between-object differences and some objects are easier to recognize in the midsegment variant (Fig. 3, see also Supplementary Fig. 1). The statistical significance of the variability across objects was assessed with a mixed-effect model that accounts for interactions between degradation and individual objects (Baayen et al., 2008; Milin, Filipovic-Durdevic, & Moscoso del Prado Martin, 2009). These interactions were highly significant (HPD 95% interval between 0.1156 and 0.1675). We will not report other analyses obtained using the mixed-effect model, since their results did not differ from the results of conventional ANOVAs.

3.2. Effect of feature type on word recognition

Would vertices play an important role in the recognition of written words – a particular set of visual shapes determined by human culture? To answer this question, we presented 6–8 letter words in midsegment and vertex forms using different levels of degradation (intact, "55% of original" and "35% of original" vari-

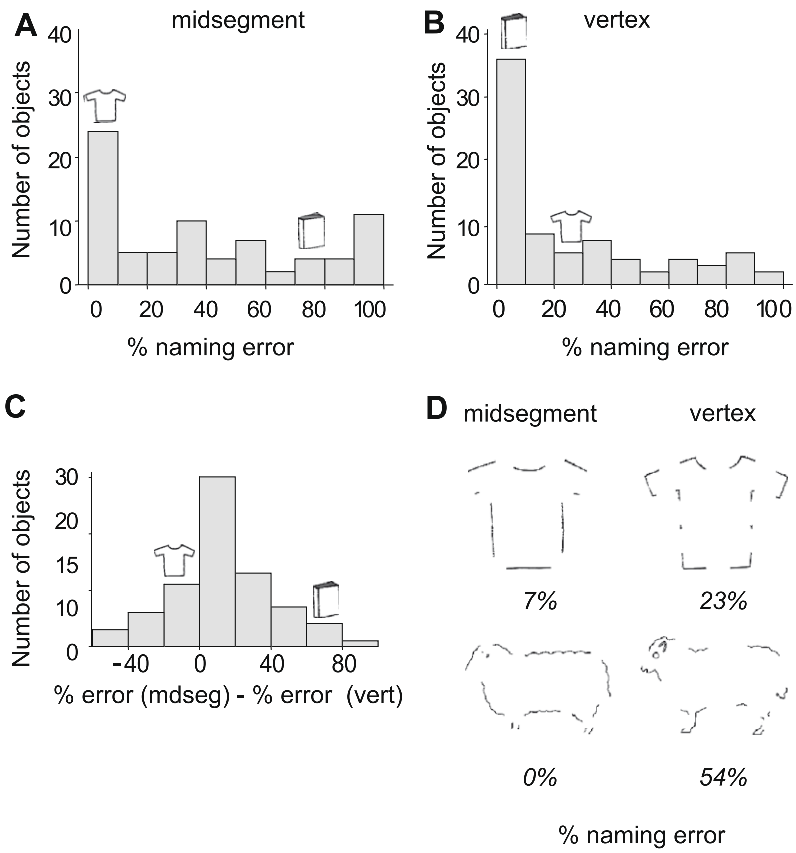

Fig. 3. Inter-object variability in naming performance. (A-B) Histogram of error rates for individual objects made out of midsegment (A) and vertex (B) features. (C) Error rate difference between the midsegment and vertex variants for each individual object: a majority of objects yield more errors when presented in midsegment form than in vertex form, but a non-negligible fraction show the opposite effect. (D) Two examples of such objects, a t-shirt and a sheep, which are recognized better in midsegment version than in vertex version. Small objects mark the positions of two examples (book and t-shirt). $n = 26$ subjects.

ants, see Fig. 1 and Section 2). In Experiment 3 we also varied the presentation times to further increase stimulus difficulty. Thus the presentation times were either 200 ms without mask or 100 ms followed by a "#######" mask.

Fig. 4 shows the naming error rates in the two experiments. The results are summarized in Table 1. As expected, in Experiment 2 (left) we found a main effect of amount of stimulus degradation, while in Experiment 3 (right) we found a main effect of exposure time and masking. More importantly, in both experiments we also found effects of feature type. For the 55% versions of the words, there was no difference in error rate between the vertex and midsegment forms. For the 35% variants, however, subjects made fewer errors for the vertex variant than for the midsegment variant. This difference was very significant both in Experiment 2, where we used only a 200 ms display time, and in Experiment 3 (Table 1). In the latter, error rates for 200 ms display were nearly identical to Experiment 2 while error rates for the short presentation time of 100 ms were naturally higher and again showed a marked advantage of the vertex features.

We analyzed reaction times for Experiment 2 (Fig. 5). The pattern was parallel to the one found for error rates. We found a main effect of degradation level ($F(2,83) = 117$, $p < .001$). With 55% stimuli, we found no difference in reaction between the vertex and midsegment variants (both 716 ms). With 35% stimuli, reaction times were slightly (14 ms) shorter for the vertex variant than for the midsegment variant, however the difference was not statistically significant ($p = 0.2$, paired t-test).

We analyzed error rates for individual words, applying the same procedure as for objects in the previous section, in order to determine whether some words were actually easier to recognize in the midsegment than in the vertex version, as was the case for some

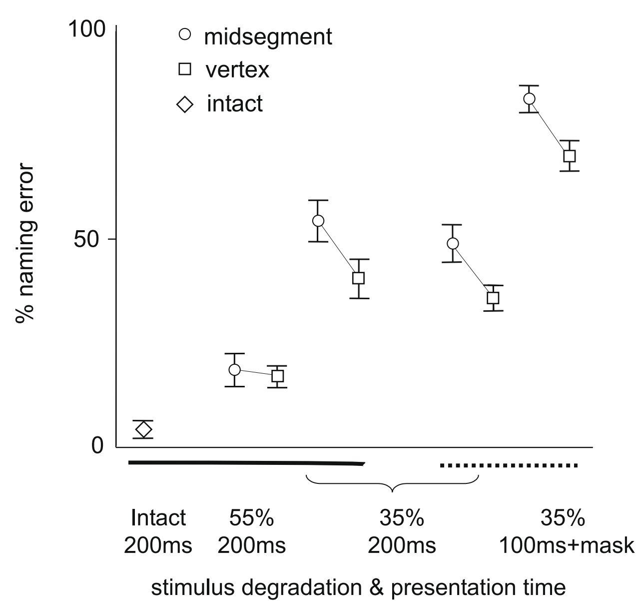

Fig. 4. Effect of feature type on word reading. We presented words in midsegment (○) and vertex (□) forms using different levels of degradation (intact, 55% of original and 35% of original, see Fig 1) and different presentation times (either 200 ms without mask or 100 ms with a "#######" mask). Results from two experiments indicate that in the most degraded stimuli, word reading is similar to object naming in that performance is worse when only the midsegment features are presented than when only the vertices are presented. Experiment 2, $n = 14$ subjects, thick solid line; experiment 3 $n = 12$ subjects, thick dotted line. Error bars denote S.E.M.

Table 1
Summary of effects of feature type on word reading.

	Experiment 2		Experiment 3					
Degradation level	55%		35%		35%		35%	
Presentation time	200 ms		200 ms		200 ms		100 ms + mask	
Stimulus variant	Vertex	Midsegment	Vertex	Midsegment	Vertex	Midsegment	Vertex	Midsegment
Percent naming error	15%	17%	37%	50%	38%	49%	73%	84%
Significance level for pairwise difference	n.s.		<.001		<.001		0.002	

In Experiment 2 we found a main effect of stimulus degradation, ($z = -14.01$, $p < .001$) and a significant ($z = 2.56$, $p = .011$) interaction between stimulus degradation and feature type. In Experiment 3 we found main effects of feature type ($z = 4.0$, $p < .001$), and exposure time + masking ($z = 10.8$, $p < .001$) with no interaction between them. Experiment 2, $n = 14$ subjects; Experiment 3 $n = 12$ subjects. Statistics were computed using a binary regression model with subjects as random covariates (see Section 2).

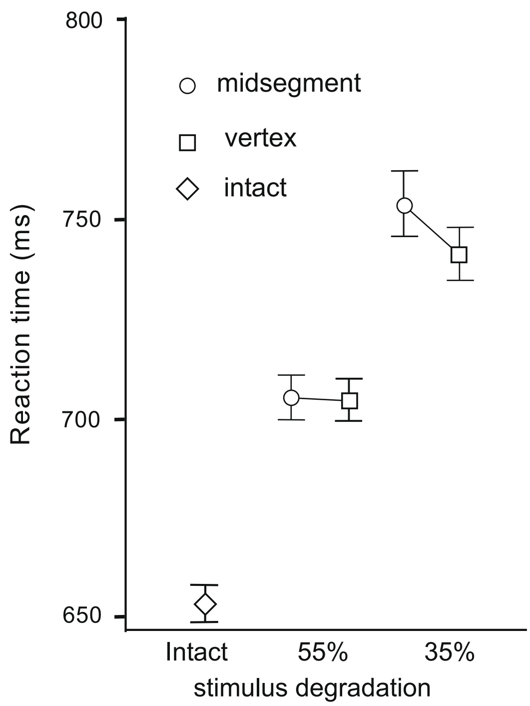

Fig. 5. Effect of feature type on word reading times. We presented words in midsegment (○) and vertex (□) forms using different levels of degradation (intact, 55% of original" (top) and "35% of original). $n = 14$ subjects. A non-significant trend for slower RTs under the midsegment presentation, parallel to the effect on error rates, is seen only at the highest level of degradation. Error bars denote S.E.M.

objects. We only considered the 35% version of words, for which the difference in error rate between the two versions was 13%, identical to the difference for objects. We found that with 23% of words, subjects performed worse with the vertex than with the midsegment version. This suggests that some words are better recognized in the midsegment version. However, this result should be interpreted with caution, since we had only 6 data points for each word, as compared to 13 for objects. Therefore, stochastic effects were stronger and most likely led us to overestimate the fraction of words better recognized in the midsegment version. The final percentage of such words would most likely be smaller if we had a more significant number of measurements for each word.

In summary, for the perception of words printed in uppercase Latin script, vertices are on the whole more important than line midsegments.

4. Discussion

4.1. Invariant features in object perception

Our first finding was that viewpoint-invariant vertex configurations are particularly informative for object perception. Objects depicted with vertices (also referred to as line junctions and line coterminations), were easier to recognize than objects depicted with line midsegments (Fig. 2). This result is similar to the result of Biederman (1987). In Biederman's paper, significant effects of feature type were found only when 35% of original picture was preserved but not for the 55% versions. They were apparent at 100 ms presentation time (>20% naming error difference), and weaker, but still present, at 200 ms presentation time (~4% naming error difference). Our experiment was more sensitive to feature type since we found a significant effect (13% naming error difference, Fig. 2A) already when 55% of original picture was present, at 200 ms presentation time. This could be either because our objects were perhaps more schematic than the objects used by Biederman (1987), or because Biederman's stimulus set included only 18 objects, which might have been relatively easier to recognize than our objects. Moreover, different viewing conditions might also explain this slight difference. Either way, Biederman needed more degradation to uncover the effect of feature type.

Our results confirm the importance of vertices in perception, but also reveal a substantial variability of this effect across objects. For 28% of objects in our dataset, subjects performed worse with the vertex than with the midsegment version – contrary to the average tendency (Fig. 3). Clearly, the visual system can use a variety of feature types besides vertices. Although the effect of deleting vertices was on average very significant, it did not abolish recognition, as would be expected if recognition was based only of vertices. Vertices are a highly informative property of real-world objects for humans (Gibson et al., 2007; Lazareva et al., 2008), animals (Gibson et al., 2007; Lazareva et al., 2008) and machines (Binford, 1981; Lowe, 1987). However, invariant properties also include other types of features such as symmetry, colinearity and curvilinearity (Biederman, 1987, Fig. 4, Lowe, 1987).

The t-shirt in Fig. 3D is an example of an object which was easier to recognize in midsegment version than in the vertex version and we can hypothesize that for that particular object, colinearities and symmetries in the trunk and the sleeves are more informative than the vertices at the corners. Stimuli easier to recognize in midsegment version included a hippo with characteristic round lines outlining its heavy body, and a sheep for which the most informative features were perhaps the wavy lines indicating its woolly hair (Fig. 3D). Deleting line midsegments might also have created misleading cues. For example, subjects often mistook the vertex hippo for a 'dog' or a 'panther'. The vertex band-aid, another object from this group was often mistaken for a 'staircase' (Supplementary Fig. 1).

To close this section we wish to note that the representation in visual cortex can be also based on features that are 1) not view invariant, but specific to particular viewpoints and 2) specific to a particular class of objects (Riesenhuber & Poggio, 2000; Ullman, 2007; Ullman, Vidal-Naquet, & Sali, 2002).

4.2. Invariant features in written word perception

Our hypothesis was that reading is based on recognition mechanisms common to all visual object recognition. We therefore expected that vertices would play an important role in the recognition of written words – although these shapes are determined solely by human culture. Indeed, we found that in a naming task subjects made fewer errors and were faster to respond for words presented in the vertex variant. This effect had the same magnitude as for objects (13% difference in naming performance). For objects, a vertex-midsegment effect appeared already at the lesser (55%) level of degradation. For words, we had to use more degraded stimuli (35% of original word remaining) to elicit a significant difference between the vertex and midsegment variants. This might be due either to the fact that letters are over-learned and overall easier to recognize (there are only 26 letters compared to thousands of possible objects), or to the fact that our objects were highly simplified, or both.

Our experimental strategy used degraded stimuli. It is known that printed words, if degraded enough, may trigger several reconstruction strategies. Thus, reading rotated, spaced, displaced (Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008) and difficult handwritten (Qiao et al., submitted for publication) words activates additional serial reading mechanisms driven by top-down influences from the parietal cortex. Similarly, partially occluded or incomplete visual stimuli activate filling-in or amodal completion processes (see Michotte, Thines, & Crabbe, 1991) that involve extensive local processing (Lamme & Roelfsema, 2000). It might be argued that the feature type effect we observed is merely due to the fact that deletions at midsegments/vertices affect reconstruction strategies in different ways. If this was the case, we would expect the difference between midsegments and vertex stimuli to disappear at a short presentation time, since these reconstruction strategies need more time to develop (at least 200 ms for filling-in Rauschenberger & Yantis, 2001). However, this was not the case. At 100 ms presentation time we observed a marked effect of feature type on naming performance (Fig. 4). Similarly, Biederman (1987) found the largest effect of vertex-midsegment difference on object recognition at 100 ms presentation time. This suggests that the vertex-midsegment effect is not due to differential activation of the above-mentioned reconstruction strategies and probably occurs at earlier stages of visual processing (Gaillard et al., in press; Lamme & Roelfsema, 2000). However, the exact time course and source of the vertex-midsegment effect remain to be determined with additional subliminal priming and/or EEG/MEG experiments.

4.3. Comparison with previous studies on letter recognition and form

The question of features used in letter perception has been addressed by several studies (Gibson, Gibson, Pick, & Osser, 1962; Grainger, Rey, & Dufau, 2008). Petit and Grainger (2002) studied the contribution of different features to perception of single letters. Their subjects performed either a letter naming task or an alphabetic decision task (decide whether the target is a letter or a non-letter). They used a masked partial priming design, where intact letter targets were preceded by subliminal, pattern-masked primes formed by deleting pixels in the target stimulus. The primes were letters in "junction" (similar to our vertex), midsegment, or "global" (dotted outline) forms. All prime forms produced significant priming effects in both tasks. However, contrary to us, they found either no difference between the effect of vertex and midsegment primes, or a bit stronger (4–12 ms, $p < .05$) priming effects for midsegment primes.

This difference between their results and ours might have several reasons. Firstly, it could be due simply to the differences in the target (words versus single letters) and manner of presentation (overt stimulus naming versus subliminal priming). Secondly, Petit and Graingers' (2002) stimuli had much lower resolution. Each of their letters was composed of only 18 pixels, while our letters had on average 125 pixels in the 55% variant and 80 pixels in the 35% variant. Each 'branch' of vertices in our stimuli was long enough to be clearly visible (Fig. 1B) while Petit and Graingers' vertices' branches were sometimes only 1 pixel long. According to our subjective perceptual experience, the 'branches' of these vertices can hardly be seen - the 'T' junctions look like triangles, for example. Consequently, vertices might not have been clear enough to activate the relevant neuronal detectors, especially in a subliminal paradigm, which may explain the discrepancy between their results and ours.

Fiset and colleagues (2008) used the bubbles technique (Gosselin & Schyns, 2001) to determine which fragments of letters are efficient for identification. Contrary to our approach this method samples all letter regions and does not manipulate *a priori* defined features, while we explicitly targeted vertices and line midsegments. Similar to us, the authors found that line vertices are important features, especially in uppercase letters. In addition, they discovered that line terminations are the most important features for letter identification (as mentioned in Section 2, we left letter terminations intact both in vertex and midsegment variants). Their results also reveal the particularities of individual letters such as the informativeness of the serif termination in the uppercase 'G' or of the descenders in lowercase 'p', 'q' 'j' and 'y' (Fig. 2 in their article). The main finding of Fiset and colleagues (2008), the importance of line terminations for letter recognition, is congruent with our concept of co-optation or "neuronal recycling" of object recognition for word reading. Line terminations are very salient stimuli for the visual system. A distinct class of primary visual cortex cells called end-stopped, or hypercomplex cells responds to terminations, i.e. they fire when a properly oriented line-end is centered in the receptive field but not when a line extends across it (Hubel & Wiesel, 1968). Thus, the detection of line terminations is a prominent property of mammalian primary visual cortices which might have been recycled for reading.

In a final analysis, Fiset and colleagues (Fig. 5 in Fiset et al., 2008) compare their human behavioral results to the performance of an artificial "ideal observer" that detects letters using all the available pixel information in the 'bubblized' letter features. The logic and method of the "ideal observer" analysis follows Pelli and colleagues' analysis on letter by letter vs. whole word recognition (2006). It compares the differences between human performance and an "ideal" benchmark to detect bottlenecks in the human system that make it sub-optimal. In contrast to Pelli, Burns, Farell, and Moore-Page (2006), however, Fiset and colleagues (2008) use "bubblized" letter features, and not letters and entire words. They found that while terminations and vertices are the most important features for human observers, they are not particularly informative for the artificial classifier. The latter 'prefers' line midsegments and curves. In other words, from an information theory point of view, the use of letter features in humans is not optimal. On the other hand, similar "ideal observer" techniques demonstrate that vertex features are highly informative in objects (Ullman, 2007; Ullman et al., 2002). As discussed further down, this fact might have interesting consequences for the hypothetical process of neuronal recycling.

4.4. Evidence for neuronal recycling in reading

The neuronal recycling hypothesis (Dehaene, 2005; Dehaene & Cohen, 2007) proposes that: (1) Visual word recognition results from recycling of a subset of neural structures used for object recognition. A telling sign of this is the location of the Visual Word

Form Area (Cohen et al., 2000; Gaillard et al., 2006) which lies next to a larger array of areas best activated by images of objects (Hasson, Levy, Behrmann, Hendler, & Malach, 2002). (2) Through functional specialization, the recycled structures acquire new computational features, e.g. the ability for case-invariant letter recognition (Dehaene et al., 2001) and sensitivity to orthographic regularities (Vinckier et al., 2007). The neuronal recycling view thus differs from propositions that reading could achieved without the need for far-reaching functional specialization (Behrmann, Nelson, & Sekuler, 1998; Price & Devlin, 2003), see reply by (Cohen & Dehaene, 2004).

The present study dealt only with the first of the above propositions, and aimed at determining whether the critical role of line vertices could be one feature of object perception preempted by the reading system. In principle, vertex recognition did not have to be recycled. One could well imagine a word recognition system that does not depend on invariant line junctions. All that it would take would be a process of learning to read that "unlearns" this particular aspect of object recognition.

However, the theoretical work of Changizi et al. (2006) argues against such far-fetched re-configuration of recognition mechanisms for reading. The authors found that in over a 100 writing systems, vertex configurations such as T or L obey a universal frequency distribution which is shared with that found in environmental images. This, in their opinion, shows that 'visual signs have been culturally selected to match the kinds of conglomeration of contours found in natural scenes', and argues in favor of biological constraints on the shape of writing systems.

Now, the experiments presented in this paper demonstrate that reading relies on the same visual features as object recognition. The minimal shapes that this system can easily represent (Tanaka, 1996) such as line junctions have been discovered and exploited in our writing systems. Thus, it is not the human cortex that has evolved for reading - there was not enough evolutionary time and pressure for such an evolution. Rather, writing systems themselves evolved under the constraint of having to remain learnable and easily recognizable by our primate visual system (Changizi et al., 2006; Dehaene, 2005; Dehaene & Cohen, 2007). For instance, no visual alphabet resembles Braille (Fig. 6A). Although Braille is an efficient binary code which is probably well-suited to the tactile system, a "visual Braille" would use poorly the capacities of our visual system. Another improbable writing system would be one based exclusively on metric properties – for example, all the letters of alphabet represented by a 'C' in different sizes and orientations: a = C, b = C, c = C, d = C, etc, (Biederman, pers. comm., Fig. 6B), This would run against the fundamental capacity of the visual system for size and orientation invariance.

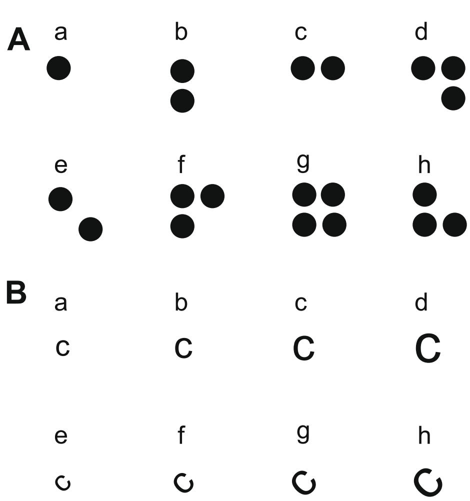

Fig. 6. "Impossible" visual alphabets. (A) Although Braille is an efficient binary code which is probably well-suited to the tactile system, a "visual Braille" would use very poorly the capacities of our visual system. (B) Another improbable writing system would be one based exclusively on metric properties – a single symbol, such as 'C' in different sizes and orientation.

As a final note we would like to highlight an apparent paradox: by virtue of their invariance for changes in viewpoint, line vertices are thought to be primarily important for the perception of 3D objects. We show that vertex encoding is recycled for reading. However, 3D invariance is not an issue for reading, because graphemes are nearly always presented on a two-dimensional surface and in a invariable orientation (e.g. left to right). Thus, in written word recognition, vertices are in fact dissociated from their original role. They are important only inasmuch as the visual system already relies on vertices for 3D object invariance, even when this function is not particularly important for reading. In fact, as mentioned above in the discussion of the study by Fiset and colleagues (Fig. 5 in Fiset et al., 2008), for an artificial "ideal observer" that uses all the information available in images of letters, vertices are not particularly informative. According to the hypothesis of neuronal recycling, vertices are particularly informative for human readers because part of the ventral visual system has evolved to recognize objects, and the optimal way for doing so was based on vertices. When this system is trained to process printed words, it keeps using vertices even though this procedure is not necessarily optimal with respect to a theoretical "ideal observer". This again shows how existing visual recognition mechanisms may have influenced the cultural evolution of writing.

Acknowledgments

We thank all subjects for their participation, Yulia Lerner for sharing images and Fabien Vinckier, Lionel Naccache, Michel Vidal-Naquet, Evelyn Eger, Jerome Sackur, Andreas Kleinschmidt, Sepideh Sadaghiani, Christophe Pallier and Rafael Malach for helpful comments. This study was supported by a grant (CORELEX) from the Agence Nationale de la Recherche. M.S. was supported by a Human Frontier Science Program fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2009.01.003.

References

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. In *Mental lexicon working papers in psycholinguistics* (pp. 1–45). Edmonton.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Behrmann, M., Nelson, J., & Sekuler, E. B. (1998). Visual complexity in letter-by-letter reading: "Pure" alexia is not pure. *Neuropsychologia, 36*(11), 1115–1132.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*, 115–147.

Biederman, I. (1995). Visual object recognition. In S. Kosslyn, & D. N. Osherson (Eds.), *An invitation to cognitive science* (Vol. 2, pp. 121–165).

Binford, T. Q. (1981). Inferring surfaces from images. *Artificial Intelligence, 17*, 205–244.

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7*(8), 880–886.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B, 272*(1560), 267–275.

Changizi, M. A., Zhang, Q., Ye, H., & Shimojo, S. (2006). The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *American Naturalist, 167*, e117–e139.

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., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., 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., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage, 40*(1), 353–366.
Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The "neuronal recycling" hypothesis. In S. Dehaene, J. R. Duhamel, M. Hauser, & G. Rizzolatti (Eds.), *From monkey brain to human brain* (pp. 133–157). Cambridge, MA: MIT Press.
Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron, 56*(2), 384–398.
Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4*(7), 752–758.
Fiset, D., Blais, C., Éthier-Majcher, C., Arguin, M., Bub, D., Gosselin, F. et al. (2008). Features for uppercase and lowercase letter identification. *Psychological Science, 19*(11), 1161–1168.
Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Baulac, M., Cohen, L., & Naccache, L. (in press). Converging intra-cranial markers of conscious access. *PLoS Biology*.
Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., Hasboun, D., et al. (2006). Direct intracranial, fMRI and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron, 50*, 191–204.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gibson, E. J., Gibson, J. J., Pick, A. D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology, 55*, 897–906.
Gibson, B. M., Lazareva, O. F., Gosselin, F., Schyns, P. G., & Wasserman, E. A. (2007). Nonaccidental properties underlie shape recognition in Mammalian and nonmammalian vision. *Current Biology, 17*(4), 336–340.
Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41*(17), 2261–2271.
Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences, 12*, 381–387.
Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34*(3), 479–490.
Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.
Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience, 23*(7), 3016–3027.
Kinzler, K. D., & Spelke, E. S. (2007). Core systems in human cognition. *Progress in Brain Research, 164*, 257–264.
Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.
Lazareva, O. F., Wasserman, E. A., & Biederman, I. (2008). Pigeons and humans are more sensitive to nonaccidental than to metric changes in visual objects. *Behavioral Processes, 77*(2), 199–209.
Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex, 12*(2), 163–177.
Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31*, 355–395.
Michotte, A., Thines, G., & Crabbe, G. (1991). *Les complements amodaux des structures perceptives. Michotte's experimental phenomenology of perception*. Hillsdale, NJ: Erlbaum.
Milin, P., Filipovic-Durdevic, D., & Moscoso del Prado Martin, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language, 60*, 50–64.
New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36*(3), 516–524.
Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.
Petit, J.-P., & Grainger, J. (2002). Masked partial priming of letter perception. *Visual Cognition, 9*, 337–353.
Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics, 9*, 127–147.
Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage, 19*(3), 473–481.
Qiao, E., Vinckier, F., Szwed, M., Naccache, L., Valabrègue, R., Dehaene, S., et al. (submitted for publication). Visual processing of handwritten words.
Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature, 410*(6826), 369–372.
Riesenhuber, M., & Poggio, T. (2000). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.
Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.
Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11*(2), 58–64.
Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7), 682–687.
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.