Interaction of top–down and bottom–up processing in the fast visual analysis of natural scenes

Delorme Arnaud[1], Rousselet Guillaume A., Macé Marc J.-M.
& Fabre-Thorpe Michèle[2]
Centre de recherche Cerveau et Cognition (UMR 5549, CNRS-UPS)
118 route de Narbonne, 31062 Toulouse, Cedex, France

[1] Delorme A., present address: Computational Neurobiology Laboratory, Salk Institute, 10010 N. Torrey Pines Road, CA92037, San Diego, USA
Tel: (858) 458-1927 x15 – Fax: (858) 587-0417
arno@salk.edu – http://www.cnl.salk.edu/~arno

[2] Corresponding author: Fabre-Thorpe, Michèle
Centre de recherche Cerveau et Cognition (UMR 5549, CNRS-UPS)
118 route de Narbonne, 31062 Toulouse, Cedex, France
Tel: (33) 5 62 17 28 07 – Fax: (33) 5 62 17 28 09
Michele.fabre-thorpe@cerco.ups-tlse.fr

Number of text pages: 17
Number of figures: 5

Abstract

The influence of task requirements on the fast visual processing of natural scenes was studied in 14 human subjects performing in alternation an "animal" categorization task and a single-photograph recognition task. Target photographs were randomly mixed with non-target images and flashed for only 20 ms. Subjects had to respond to targets within 1 s. Processing time for image-recognition was 30-40 ms shorter than for the categorization task, both for the fastest behavioral responses and for the latency at which event related potentials evoked by target and non-target stimuli started to diverge. The faster processing in image-recognition is shown to be due to the use of low-level cues, but source analysis produced evidence that, regardless of the task, the dipoles accounting for the differential activity had the same localization and orientation in the occipito-temporal cortex. We suggest that both tasks involve the same visual pathway and the same decisional brain area but because of the total predictability of the target in image-recognition, the first wave of bottom-up feed-forward information is speeded up by top down influences that might originate in the prefrontal cortex and preset lower levels of the visual pathway to the known target features.

1. INTRODUCTION

Spotting a specific object among others is an every day task that appears trivial but raises a number of questions concerning the underlying visual processing. In visual search tasks, subjects are asked to look for a pre-specified target embedded in distractor arrays. Typically, for low-level features, ERP studies suggest that a visual decision can be made in about 150 ms [1,21,34]. This latency increases when targets are defined by a conjunction of characteristics such as form and color [18], although pop out has been reported for some specific conjunction of low-level features [7,21,28,38]. Surprisingly, 150 ms has also been reported to be the minimal processing time to differentiate between different classes of natural images. Using a superordinate categorization task in which human subjects had to respond when a natural image that they had never seen before contained an animal, Thorpe et al. [36] showed that visual evoked potentials recorded on correct target trials differed sharply from those recorded on correct distractor trials at about 150 ms after stimulus onset. This differential brain activity has been found at the same latency with non-biological relevant categories of objects such as "means of transport" and has been shown to be related to "visual decision making" rather than physical differences between photographs belonging to different categories [40]. This speed of processing could well be seen for any well-learned object-category [32]. In such categorization tasks, very different objects have to be grouped together (i.e. a snake and a flock of sheep) and performance cannot rely on the analysis of a single low-level cue or even on a single conjunction of low-level cues. When considering this very short delay together with the anatomy and physiology of the visual system, it was argued that such severe temporal time constraints imply that the underlying processing probably relies on feed-forward mechanisms during a first wave of visual information [35,36].

It thus seems that high-level search tasks such as looking for an animal in a natural scene might be performed as fast as the simplest pop-out search tasks. To explain speed of processing in visual search tasks, emphasis had been put on the target saliency, and on the number of diagnostic stimulus features [33]. However, increasing stimulus diagnosticity in the animal categorization task of natural images by using highly familiar photographs failed to induce a decrease of the minimal processing time: subjects could categorize novel images as fast as images on which they had been extensively trained [8].

Thus, the fast visual processing mode that underlies rapid-categorization cannot be speeded up when top-down pre-setting of the visual system is optimized with experience. However, it is a difficult experimental issue to determine the relative importance of bottom-up and top-down processes. To investigate further how top-down knowledge related to task requirements could influence the visual analysis of natural images, we tested human subjects in a task in which they were assigned a given photograph as target and had to detect this single target-photograph among a variety of different non-target stimuli. Being fully briefed about the target should allow subjects to maximize the use of top-down influences and to rely only on a limited number of low-level cues specific to the target-image.

In the present experiment, we studied the fast processing of natural images in human subjects performing in alternation the superordinate "animal / non-animal" categorization task and the single-photograph recognition task. Along with behavioral performance, analysis involved associated ERPs and localization of brain sources to investigate the neural dynamics of early information processing. Since both tasks used the same natural images as stimuli and required the same motor response, any processing differences should be related to task requirements.

2. METHODS

4

Stimuli

All stimuli used in the two tasks were photographs of natural scenes (Corel CD-ROM library). In each group, images were chosen to be as varied as possible (Figure 1). Subjects were tested on blocks of 100 stimuli including 50 % targets and 50 % distractors. In the categorization task 1000 photographs were used (50 % distractors and 50 % targets) and each of them was seen only once by each subject. The target-photographs included pictures of mammals, birds, fish, arthropods, and reptiles. There was no a priori information about the size, position or number of targets in the photograph. There was also a wide range of non-target images, with outdoor and indoor scenes, natural landscapes or city scenes, pictures of food, fruits, vegetables, trees and flowers....

In the recognition task, as in the categorization task, targets and non-targets were equiprobable in each block of 100 images so that the target-photograph assigned to a given block was seen 50 times among 50 varied non-target photographs that did not contain an animal. Each of the 14 subjects was tested with 15 targets (a total of 210 targets) and the same 750 non-target stimuli. In the 210 photographs used as targets, 140 (10 images per subject) contained an animal and were thus similar to the target photographs used in the categorization task. They had been categorized by human subjects in a previous study [8] and were known to offer different levels of difficulty. The remaining 70 (5 images per subject) did not contain any animal and were thus homogenous with the non-targets used in both tasks.

Task and protocol

Fourteen human subjects (7 women and 7 men, mean age 26 ranging from 22 to 46), with normal or corrected to normal vision volunteered for this study. Participants sat in a dimly lit room at 110 cm from a color computer screen piloted from a PC computer. They were required to start a block of 100 images

5

by pressing a touch-sensitive button. A small fixation point (< .1° of visual angle) appeared in the middle of the black screen. Then, an 8-bit color vertical photograph (256 pixels wide by 384 pixels high which roughly correspond to 4.5° X 6.5° of visual angle) was flashed for 20 ms using a programmable graphic board (VSG 2.1, Cambridge Research Systems). The short presentation time prevented any exploratory eye movement. The stimulus onset asynchrony (i.e. time between the onset of one image and the onset of the next image in a series) was random between 1800 ms and 2200 ms.

Subjects had to give a go/no-go response: releasing the button as quickly and accurately as possible when they saw a target-image but keeping their finger(s) on the button on non-target trials. They were given a maximum of 1000 ms to respond, after which delay any response was considered as a no-go response.

On two different days, subjects were tested on 10 categorization blocks and 10 recognition blocks, alternating between the two tasks within a session while their associated EEG was recorded. In the animal categorization task, subjects had to respond whenever the picture contained an animal. In the target-image recognition task, a given animal image was assigned as the target for the following block of 100 images. The 5 image-recognition control blocks using images that did not contain an animal were inserted at regular intervals.

For the image-recognition task, each testing block was preceded by a learning phase during which the subject was presented with the target-photograph which was both repeatedly flashed for 20 ms (similar to the testing conditions) and presented for 1000 ms to allow ocular exploration (3*5 flashes intermixed with 2 long -1000 ms- presentations). Participants were instructed to carefully inspect and memorize the target-image in order to respond to it in the following sequence of images as fast and as precisely as possible. The testing block started immediately after the learning phase.

Evoked–Potential Recording and Analysis

Electric brain potentials were recorded from 32 electrodes mounted on an elastic cap (Electro-cap International Inc). Data acquisition was made at 1000 Hz using a SynAmps recording system (Neuroscan Inc.) coupled with a PC computer. The analog low-pass filter was set at 500 Hz and the default SynAmps analog 50-Hz notch filter was used. Impedances were kept below 5 kOhms. Potentials were recorded with respect to common reference Cz, then average re-referenced. Potentials on each trial were baseline corrected using the signal during the 100 ms that preceded the onset of the stimulus. Trials were checked for artifacts and discarded using a [-50; +50 μV] criterion over the interval [-100; +400 ms] at frontal electrodes for eye movements and a [-30; +30 μV] criterion on the period [-100; +100 ms] at parietal electrodes to discard alpha brain waves. Only correct trials were considered for ERP averages. The waveforms were low-pass filtered at 35 Hz for use in graphics. Inter-subject two-tailed statistical *t*-tests (13 degrees of freedom) were performed on unfiltered ERPs for each electrode to evaluate the latency at which target ERPs diverged from non-target ERPs. This differential activity onset was defined as the time from which 15 consecutive values were statistically different to compensate for multiple comparisons. We computed significance for all electrodes but focused on two groups: frontal electrodes (10-20 system nomenclature: Fz, FP1, FP2, F3, F4, F7, F8) and occipital electrodes (10-20 system nomenclature: O1 & O2 with the addition of Oz, I, O1', O2', PO9, PO10, PO9', PO10') where the differential activity reached the highest amplitude. The additional occipital electrodes have the following spherical coordinates (theta/phi): Oz = 92/-90, I = 115/-90, O1′ = -92/54, O2′ = 92/-54, PO9 = -115/54, PO10 = 115/-54, PO9′ = -115/72, PO10′ = 115/-72.

Source localization

The source analysis was performed using a 4-shell ellipsoidal model and using BESA (Brain Electrical Source Analysis, version 99). Because of temporal muscle contraction, the two most temporal electrodes were too noisy and were discarded from the analysis. All other electrodes were used to localize the equivalent dipoles. Grand-average waveforms were low-pass filtered at 35 Hz before analysis. Pairs of dipoles were placed in a central position, given a spatial symmetry constraint, then fitted in location and orientation for a particular time window (simplex algorithm).

3.RESULTS

The aim of this study was to compare the visual processing of a natural image when the task requirements called for the representation of a high-level object category such as "animal" or when it could be performed using short-term memory of low-level cue(s). Behavior and ERPs were recorded and analyzed in all subjects.

Behavioral results: recognition vs. categorization

The analysis of behavioral performance included accuracy, speed of response and a study of the non-target images that incorrectly induced a go-response.

Accuracy. Although extremely good in both tasks (93.1 % correct in the categorization task; 98.7 % correct in the recognition task) accuracy was significantly better in the recognition task (two-tailed $\chi2$: df =1, $p < .0001$), an effect that was found to be significant at $p < .05$ for each individual subject. An accuracy bias was found in both tasks, but whereas this bias was in favor of correct no-go responses in the categorization task it was in favor of correct go responses in the recognition task. Thus, subjects were slightly better at ignoring distractors than responding to animal-targets in the categorization task

(93.9 % vs. 92.4 %; two-tailed $\chi2$: *df* =1, *p* < .0001) whereas they were more accurate at detecting the target-image in the recognition task than at ignoring non-target images (99.7 % vs. 97.5 %; two-tailed $\chi2$: *df* =1, *p* < .0001). This result provides an argument for the use of different strategies in the 2 tasks that will be discussed later.

Reaction time (RT). As illustrated in Figure 2, reaction times were significantly faster for the recognition task (median RT: 337 ms) than for the categorization task (median RT: 400 ms; two-tailed Mann Whitney U test: *p* < .0001). For individual subjects this difference was always significant (*p* < .01).

Processing speed can be measured using median RT or mean RT, but these values do not reflect all aspects of processing speed. One very useful value is the minimal processing time needed to complete the tasks. The average slower speed in the categorization task could be due to some difficult photographs that need longer processing time [8]. Thus, although the average processing time could be shorter in the recognition task, the minimal processing time might be similar in both tasks. As in our experimental protocol targets and non-targets were equiprobable in both tasks, we defined the minimal processing time (Figure 2) as the first time bin for which correct hits to targets started to significantly outnumber false alarms to non-targets. Responses triggered with shorter latency but with no bias towards correct go-responses were presumably anticipations initiated before stimulus processing was completed. Using 10-ms time bins, this "minimal processing time" was found significant at 220 ms (two-tailed $\chi2$: *df* =1, *p* < .0001) in the recognition task and at 260 ms in the categorization task (two-tailed $\chi2$: *df* =1, p = .0007). The minimal processing time to reach decision was thus shortened by about 40 ms in the recognition task relatively to the categorization task. However, this shortening of RT latencies can be seen in Figure 2 as a shift of the entire RT distribution of the recognition task toward shorter latencies, from the earliest to the latest behavioral responses.

9

Control set. The results obtained in the recognition task with the control sets (that used non-animal target pictures) show again the better accuracy and the shorter processing time associated with tasks that only require image recognition (Figure 5). Subjects scored 98.3 % correct, with a median RT for correct go-responses at 348 ms. These scores are slightly below the performance level observed when the one-image target contained an animal (respectively 98.7 % and 337 ms), a result that could be due to higher similarities with the distractors, but the minimal processing time was found at exactly the same latency (220 ms) in both cases (p <.0001, $\chi2$ test evaluated over every 10 ms time bin).

Errors. A question that needs to be raised concerns the kind of errors that are produced in both tasks. In the categorization task, false alarms on distractors were slightly less common than target misses, and so far it has rarely been possible to objectively determine the reasons for these errors. In contrast, the errors produced in the recognition task were often seen with non-target images that share some obvious low-level properties with the memorized target image. These features (Figure 1) appear to be related to coarse orientation of objects, prevailing color, patches of color(s) in a given location, context or object identity, spatial layout or complexity of the scene... When performing the recognition task, subjects were thus relying on low-level visual cue(s) that could differ from one memorized target to another.

Event-related potentials

ERPs were considered separately for correct target and correct distractor trials (Figure 3). Using both individual data and grand average ERPs, the differential brain activity between the two types of trial was assessed in the two tasks by subtracting the average ERP on correct distractor trials from the average ERP on correct target trials. It is commonly assumed that the averaged electrical responses recorded from the scalp result from stimulus-

10

evoked brain events and that the amplitude and latency of the various components of this evoked response reflect the most relevant features of the brain processing dynamics. Recently it has been shown [23] that these deflections might be generated by partial stimulus-induced phase resetting of multiple electroencephalographic processes. However, by using the difference between the two ERPs, no assumption is made about the relevance of the different ERP components, since the question that is addressed concerns only the differences in the cerebral processing of targets and distractors. The onset latency of this differential activity –that might correspond to the minimal visual processing time to differentiate a target from a distractor- was assessed using a two-tailed paired t-test performed for each 1 ms time bin and for each electrode (see Methods).

As reported in previous studies using this categorization task, a positive differential activity, was clearly seen on frontal electrodes [8,36]. On occipital sites, a mirror differential activity of inverse polarity was observed [10]. The results are illustrated on Figure 3 and show that ERPs to targets and non-targets superimposed very well until about 170 ms at which point they diverged abruptly (two-tailed paired t-test: df =13, p < .02; occipital: 169 ms; frontal: 179 ms).

In the recognition task, the ERPs on correct target trials were computed separately for the two different sets of target-images (animal and control non-animal sets) and for their associated non-target images (Figure 5, B). The grand average ERPs computed on all the non-targets superimposed perfectly (Figure 5, B, middle traces) showing that there was no bias in the high variety of distractors used with the two different target sets. On the other hand, ERPs averaged separately on correct trials for the two target sets showed some differences (Figure 5, B, upper traces). The onset latency of the differential ERPs (Figure 5B, lower traces) was found at 135 ms in the animal picture recognition task (two tailed paired t-test : df =13, p < .02; occipital : 135 ms;

11

frontal 148 ms), a latency virtually identical to the one found in the non-animal picture recognition task (two tailed paired t-test : df =13, p < .02; occipital : 134 ms; frontal : 145 ms). Although the onsets were similar for these two sets of recognition targets, they diverged shortly after, the amplitude of the differential ERP increasing with a steeper slope with animal pictures targets. However, in the two sets of target-images, the computed differential activities reached similar amplitudes (on FZ electrode, animal pictures: 5.5 µV; non-animal pictures: 5.1 µV); but, the peak amplitude was observed earlier with animal images (233 ms) than with the set of non-animal images (255 ms). These differences at the ERP level might reflect the higher diagnosticity of animal images among non-animal images compared to the recognition of non-animal images among similar pictures.

Thus, in the picture recognition task, a clear differential activity was also observed at all sites but its onset was seen around 140 ms, much earlier than in the categorization task regardless of whether the images contained an animal or not. Consistent with this result, the difference between the two tasks also reached significance at about 140 ms (two-tailed paired t-test: df =13, p < .02; occipital 141 ms; frontal 158 ms). Thus differential activity between target and non-target trials developed much earlier and reached a much higher amplitude in the recognition task than in the categorization task (5.3 µV vs. 2.9 µV for electrode Fz). Moreover, the peak of amplitude was observed at similar latencies in both tasks when pictures contained an animal (animal categorization: 234 ms, image-recognition: 235 ms).

In both tasks the differential ERP between animal-target and non-target ERPs also showed an early small deflection that reached significance at about the same latency in the categorization task (two-tailed paired t-test: df =13, p < .02; first occipital electrode: 98 ms; first frontal electrode 120 ms) and in the recognition task (two tailed paired t-test: df =13, p < .05; occipital: 100ms;

frontal: 112 ms). This small deflection does not appear with non-animal target images in the recognition task (Figure 5B, lower traces) and might thus be linked to statistical differences in physical properties of different subsets of images as documented recently [40].

Source localization and activation dynamics

For both tasks we used an ellipsoidal source model in the software BESA to analyze the dipole source localization of the differential ERP waveforms and the time course of their activities (Figure 4). Despite the strong constraints imposed on the model (large time window of 80 ms and only 2 dipoles that were required to be symmetrically positioned), residual variance was kept under 4 % for both tasks (residual variance: 3.9 % in the categorization task and 2.2 % in the recognition task), as already found in other studies using the categorization task [10]. Models using shorter and different time windows produced dipole localization that could not be distinguished from those illustrated in Figure 4. Thus, most of the difference between ERPs to target and non-targets can be explained by a single bilaterally activated brain area located ventrally and laterally in the occipital lobe, in a region that probably corresponds to extra-striate visual cortex. The localization and orientation of the dipoles were similar for the two tasks, the most obvious difference between the observed scalp signals being the time-course of the differential activity which started earlier in the recognition task.

In the recognition task, the two sets of images were analyzed separately and were found to be associated with non-distinguishable dipoles that accounted in both cases for about 98% of the differential ERP waveforms. The only difference was seen in the temporal dynamics of activation of both pairs of dipoles that were associated with a stronger activity increase from 150 ms onwards with the set of animal targets, reaching earlier its maximal amplitude.

13

4. DISCUSSION

The results of the present study show that the processing time of natural scenes by the human visual system depends on task instructions. When subjects are required to recognize a given target-image, they can rely on a variety of low-level cues, a hypothesis supported by the high similarity between the target and the non-target scenes that induced response errors. Consequently, the subjects were faster and more accurate in this natural scene recognition task than when they categorized the same type of natural images on the basis of the presence of an animal, a task that presumably requires access to more abstract representations. The results also provide some evidence that regardless of the visual analysis required in either task, the perceptual decision is made in the same brain structure and the visual information probably processed along the same visual pathway.

The visual processing required for recognizing a given target-image is done in a delay that is about 30-40 ms shorter than the visual analysis required for detecting an animal in the same image. This delay is observed for both the latency at which the earliest behavioral responses are produced and the onset latency of the differential cerebral activity (used as an index of the perceptual decision). It increases to 60 ms when considering the median reaction time, reflecting the fact that the variation in response latencies is larger in the animal categorization task than in the image-recognition task (Figure 2) because of a larger difficulty range in the categorization task.

One could argue that the main difference between the two tasks is due to a novel vs. familiarity effect. Whereas the categorization task is exclusively performed with previously unseen images (trial unique presentations), the target-image recognition task involves the repetitive visual processing of a recently memorized photograph (i.e. "familiar") among non-target images that have never been seen before (i.e. "novel"). Indeed, it has been shown using event-related fMRI, that the activity of brain areas that are thought to be

14

involved in scene categorization (extrastriate visual cortex, inferotemporal cortex and prefrontal cortex) is modulated by stimulus repetition in subjects performing a rapid classification of pictures [4]. However, in the "animal" categorization task, we have recently shown that extensive experience with a given set of natural scenes did not result in faster behavioral responses than with completely novel images nor reduce the latency of the differential ERPs [8]. In agreement with other ERP studies using words, faces and other visual stimuli [12,22,31,39], familiarity effects were not seen until about 300-360 ms post-stimulus and thus could not account for speeding up the visual processing in the recognition task used here.

Various interpretations could account for our results. As target-image recognition task relies on detection of low-level cues, one possibility is that the faster analysis could simply result from the by-pass of higher processing stages that would only be necessary to reach a decision in the superordinate "animal" categorization task, when access to abstract representations is specifically required. In the recognition task, the perceptual decision could be made in brain structures considered as lower in the hierarchy of visual processing but in which low-level features would be already fully analyzed and accessible. Decisions could be made in area V4 or even in the primary visual cortex V1 as suggested by Barbur et al. [2]. Alternatively we would like to argue that visual information is analyzed along the same brain pathway [16] but that the higher target predictability in the image-recognition task allows faster processing of the pertinent cues using top-down connections to preset neuronal assemblies at various levels of the visual pathway.

The main result supporting this alternative view is the location of the dipoles accounting for 96 % and more of the differential activity recorded in both tasks. Even though the 32-recording-site set-up and the ability of the BESA software to specify accurately the "absolute" location of the brain activity may be questioned, the fact that, regardless of the task, the dipoles

15

were found at very similar positions and orientations in the brain appears difficult to explain if the underlying brain areas were not the same. In both tasks, the perceptual decision could therefore involve the same cerebral structures, most probably the occipito-temporal visual areas involved in object recognition. The location has been confirmed using the same categorization task with an event-related fMRI study [9], and found to be close to areas such as the fusiform gyrus involved in the recognition of various stimuli such as faces, objects or animals [5,14,20]. In correlation with the differential activity that develops 30-40 ms earlier in the target-image task, the main difference between our two tasks was found in the temporal dynamics and amplitude of the dipole activation (Figure 4) that developed earlier and reached higher amplitude in the image-recognition task.

In preceding studies using the animal categorization task we have already argued that the short latency at which the scalp differential activity starts to develop imposes such a high temporal constraint that the perceptual decision presumably relies essentially on feed-forward processing [8,35,36]. We postulated that information from the retina had to reach the primary visual cortex, area V1 (via the thalamus), and was subject to further processing in areas V2 and V4 before reaching the high-level brain areas involved in object recognition. These various processing steps are likely to be just as essential in the target-image recognition. Thus the most likely interpretation still relies on a faster visual processing of these images because of total target predictability.

In both tasks, speed of bottom-up processing would depend upon the tuning of neuronal populations along the visual pathways and thus on stimulus diagnosticity. Such bias has been shown for spatial frequencies [29], suggesting that a given scene might be flexibly encoded and perceived at the scale that optimizes information for the on-going task. Automatic target priming has been shown for color and spatial position in pop-out tasks [24,25] and has been attributed to temporary representations that could be updated on the basis of

16

task demands. Saccade latency can be shortened by 30 ms and more, an effect linked to diagnosticity since it builds up with target color repetition [26]. In our tasks, we would expect top-down influences to bias bottom-up visual processing more heavily and more precisely in the recognition task than in the categorization task. The recognition of a target scene might be achieved using a carefully chosen low-level feature or a simple combination of characteristics (a blob of a given color or orientation for example). Compatibility would be maximal in this task because every target-image would activate all preset neuronal populations. Moreover, as the specific location of this feature in the image is also known, focalized spatial attention could be allocated at the exact location of the screen where the cue is going to appear when the target is flashed; a view that is supported by our analysis of the images that induced false alarms. In contrast, in the categorization task, the subject needs to process evenly the whole natural scene: the location of the target-animal in the photograph is unknown and although many features (an eye, a paw, a tail, a beak, a wing...) are diagnostic of the presence of an animal, none of them is necessary to classify an image as a target. Thus the presetting of the visual system cannot be as highly specific as in the recognition task and could not rely on the same features. Indeed, whereas color appears as an important diagnostic feature in the image-recognition task, we have shown that the fast responses in the "animal" categorization task do not rely on color cues [6]. A strong modulation of color processing could be due to top-down influences from high-level predictions about color-specific features [19].

Among the brain structures that might heavily influence the visual pathway through descending connections depending on behavioral requirements is the prefrontal cortex [3,27]. In a categorization task, the firing of prefrontal neurons reflects category membership rather than simple processing of the physical characteristics of the stimuli [11]. In the target-image recognition task, the activity in the frontal cortex is probably very similar to that recorded in a

delayed matching to sample task with elevated activity during delay periods [13,15]. Moreover, prefrontal neurons can also convey information about both the physical characteristics of a stimulus and its location [30], a combination of cues used in the target-image recognition task. Thus, in the target-image recognition task, prefrontal activity could very precisely modulate the neuronal activity along the visual pathway [17] to optimize, for each memorized target, the processing of the selected pertinent cues.

Whereas total predictability speeds up visual processing, we showed using a control set of target images that presetting does not have the same strength for all natural scenes. Scenes with animals were, on average, recognized faster than scenes without animals. Certainly some features might be more salient in animal photographs presented among non-animal photographs, whereas the control set of non-animal images presented among other non-animal pictures could lack this diagnostic advantage. Another possible explanation may lie in the performance, in alternation, of the animal categorization task and the image recognition. Subject might have difficulty in inhibiting totally the presetting of neuronal populations tuned to animal features.

Another point that needs stressing is the fact that, in our preceding studies, the onset of the differential activity was found at about 150 ms for the categorization whereas in the present study it was found about 20-30 ms later. Image size or presentation cannot account for this increased onset latency. On the other hand, this difference could be explained by the switching between two different tasks that required different presettings of the visual system as it has also been seen in another experimental protocol using two different interleaved tasks (manuscript in preparation). It might be that, had we used a blocked procedure in which subject would have completed all the testing series of one task before completing the second task we would have ended with even shorter differential activities.

18

Regardless of the task, we suggest that natural images are processed along the same visual circuit and that a perceptual decision is made in the same brain area but that the processing speed of bottom-up information is highly dependent upon the subject expectancy and the strength of top-down influences. However, we evaluated the temporal cost of the higher-level visual computations needed to perform the superordinate "animal" categorization task at about 30-40 ms. This temporal cost appears low when considering the discrepancy in task requirements. The answer might be in the level of complexity of the most informative features for classification. Fast super-ordinate categorization might rely on diagnostic features of intermediate complexity [37], accessible with coarse visual information rather than on fully integrated high-level object representations.

References

[1] Anllo-Vento, L., Luck, S.J. and Hillyard, S.A., Spatio-temporal dynamics of attention to color: evidence from human electrophysiology, *Hum Brain Mapp*, 6 (1998) 216-38.

[2] Barbur, J.L., Wolf, J. and Lennie, P., Visual processing levels revealed by response latencies to changes in different visual attributes, *Proc R Soc Lond B Biol Sci*, 265 (1998) 2321-5.

[3] Barcelo, F., Suwazono, S. and Knight, R.T., Prefrontal modulation of visual processing in humans, *Nat Neurosci*, 3 (2000) 399-403.

[4] Buckner, R.L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B. and Dale, A.M., Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI, *Neuron*, 20 (1998) 285-96.

[5] Chao, L.L., Haxby, J.V. and Martin, A., Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects, *Nat Neurosci*, 2 (1999) 913-9.

[6] Delorme, A., Richard, G. and Fabre-Thorpe, M., Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans, *Vision Res*, 40 (2000) 2187-200.

[7] Enns, J.T. and Rensink, R.A., Influence of scene-based properties on visual search, *Science*, 247 (1990) 721-3.

[8] Fabre-Thorpe, M., Delorme, A., Marlot, C. and Thorpe, S.J., A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes, *J Cog Neurosci*, 13 (2001) 171-180.

[9] Fize, D., Boulanouar, K., Chatel, Y., Ranjeva, J.P., Fabre-Thorpe, M. and Thorpe, S.J., Brain areas involved in rapid categorization of natural images: an event-related fMRI study, *Neuroimage*, 11 (2000) 634-43.

[10] Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B. and Thorpe, S.J., Rapid categorisation of foveal and extrafoveal natural images: Associated ERPs and effect of lateralisation, *(Submitted)*.

[11] Freedman, D.J., Riesenhuber, M., Poggio, T. and Miller, E.K., Categorical representation of visual stimuli in the primate prefrontal cortex, *Science*, 291 (2001) 312-6.

[12] Friedman, D., Cognitive event-related potential components during continuous recognition memory for pictures, *Psychophysiol*, 27 (1990) 136-48.

[13] Fuster, J.M. and Alexander, G.E., Neuron activity related to short-term memory, *Science*, 173 (1971) 652-4.

[14] Gauthier, I., Skudlarski, P., Gore, J.C. and Anderson, A.W., Expertise for cars and birds recruits brain areas involved in face recognition, *Nat Neurosci*, 3 (2000) 191-7.

[15] Goldman-Rakic, P.S., Cellular basis of working memory, *Neuron*, 14 (1995) 477-85.

[16] Grill-Spector, K. and Kanwisher, N., Different recognition tasks activate a common set of object processing areas in the human brain, *Soc Neurosci Abstr* (2000) 686.6.

[17] Hasegawa, I. and Miyashita, Y., Categorizing the world: expert neurons look into key features, *Nat Neurosci*, 5 (2002) 90-1.

[18] Hillyard, S.A. and Anllo-Vento, L., Event-related brain potentials in the study of visual selective attention, *Proc Natl Acad Sci U S A*, 95 (1998) 781-7.

[19] Hopf, J.M., Vogel, E., Woodman, G., Heinze, H.J. and Luck, S.J., Localizing visual discrimination processes in time and space, *J Neurophysiol*, 88 (2002) 2088-95.

[20] Kanwisher, N., McDermott, J. and Chun, M.M., The fusiforme face area: a module in human extrastriate cortex specialized for face perception, *J Neurosci*, 17 (1997) 4302-11.

[21] Karayanidis, F. and Michie, P.T., Evidence of visual processing negativity with attention to orientation and color in central space, *Electroencephalogr Clin Neurophysiol*, 103 (1997) 282-97.

[22] Liu, T. and Cooper, L.A., The influence of task requirements on priming in object decision and matching, *Mem Cognit*, 29 (2001) 874-82.

[23] Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E. and Sejnowski, T.J., Dynamic brain sources of visual evoked responses, *Science*, 295 (2002) 690-4.

[24] Maljkovic, V. and Nakayama, K., Priming of pop-out: I. Role of features, *Mem Cognit*, 22 (1994) 657-72.

[25] Maljkovic, V. and Nakayama, K., Priming of pop-out: II. The role of position, *Percept Psychophys*, 58 (1996) 977-91.

[26] McPeek, R.M., Maljkovic, V. and Nakayama, K., Saccades require focal attention and are facilitated by a short-term memory system, *Vision Res*, 39 (1999) 1555-66.

[27] Miller, E.K. and Cohen, J.D., An integrative theory of prefrontal cortex function, *Annu Rev Neurosci*, 24 (2001) 167-202.

[28] Nakayama, K. and Silverman, G.H., Serial and parallel processing of visual feature conjunctions, *Nature*, 320 (1986) 264-5.

[29] Oliva, A. and Schyns, P.G., Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli, *Cognit Psychol*, 34 (1997) 72-107.

[30] Rainer, G., Asaad, W.F. and Miller, E.K., Memory fields of neurons in the primate prefrontal cortex, *Proc Natl Acad Sci U S A*, 95 (1998) 15008-13.

[31] Rugg, M.D., Soardi, M. and Doyle, M.C., Modulation of event-related potentials by the repetition of drawings of novel objects, *Brain Res Cogn Brain Res*, 3 (1995) 17-24.

[32] Schendan, H.E., Ganis, G. and Kutas, M., Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms, *Psychophysiol*, 35 (1998) 240-51.

[33] Schyns, P.G., Diagnostic recognition: task constraints, object information, and their interactions. In M.J. Tarr and H.H. Bülthoff (Eds.), *Object recognition in man, monkey, and machine*, Elsevier Science Publishers, Amsterdam, 1998, pp. 147-179.

[34] Sugita, Y., Electrophysiological correlates of visual search asymmetry in humans, *Neuroreport*, 6 (1995) 1693-6.

[35] Thorpe, S.J. and Fabre-Thorpe, M., Seeking categories in the brain, *Science*, 291 (2001) 260-3.

[36] Thorpe, S.J., Fize, D. and Marlot, C., Speed of processing in the human visual system, *Nature*, 381 (1996) 520-2.

[37] Ullman, S., Vidal-Naquet, M. and Sali, E., Visual features of intermediate complexity and their use in classification, *Nat Neurosci*, 5 (2002) 682-7.

[38] Valdes-Sosa, M., Bobes, M.A., Rodriguez, V. and Pinilla, T., Switching attention without shifting the spotlight object-based attentional modulation of brain potentials, *J Cogn Neurosci*, 10 (1998) 137-51.

[39] Van Petten, C. and Senkfor, A.J., Memory for words and novel visual patterns: repetition, recognition, and encoding effects in the event-related brain potential, *Psychophysiol*, 33 (1996) 491-506.

[40] VanRullen, R. and Thorpe, S.J., The time course of visual processing: from early perception to decision- making, *J Cogn Neurosci*, 13 (2001) 454-61.
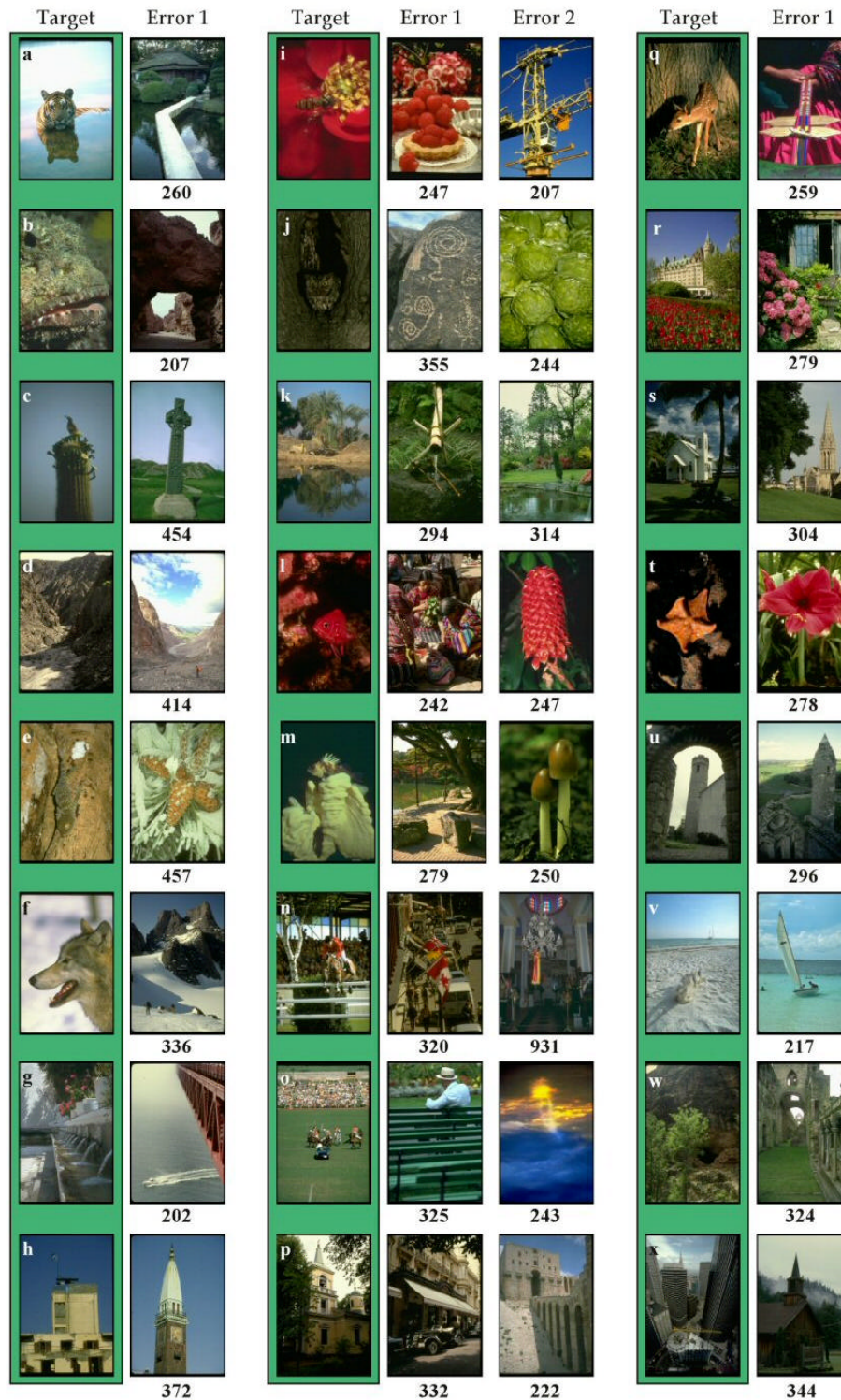
Figure 1: Delorme et al.

Figure 1. Targets and associated errors in the recognition task. Target-images used in the recognition task are illustrated on a green background. The figures show the high variety of the animal images used in the 10 testing blocks (images a, b, c, e, f, i, j, k, l, m, n, o, q, v) in which animals are sometimes hardly visible (e, i, j, v) and the non-animal images used in the 5 control blocks (images d, g, h, p, r, s, t, u, w, x). On the right of each target-image is shown the non-target photograph(s) that induced a false alarm. Errors can clearly be related to global orientation (a, c, d, g, h…), color (e, i, j, l…), color patches in specific locations (n, t…), object identity or semantics (p, s, x…), spatial layout of the scene (b, e, f, k, n, v….) or any combination. The figures below each error indicate the reaction time of the incorrect go response. Similar natural images were used in the categorization task.
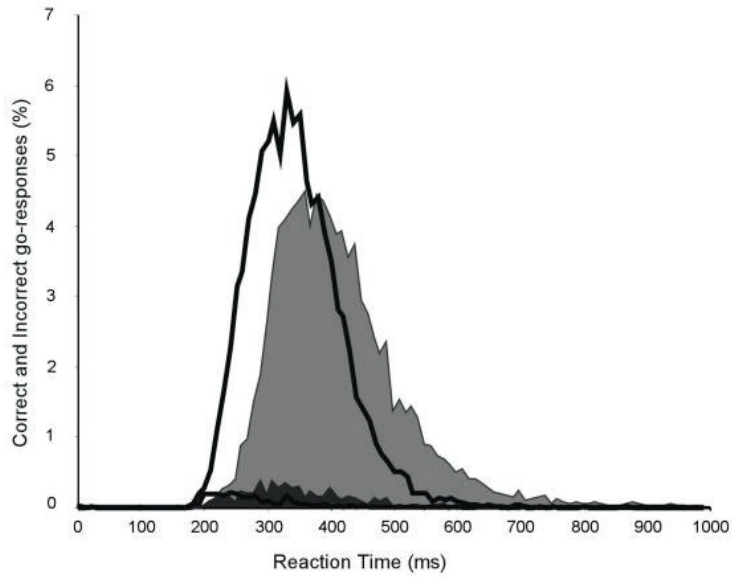
Figure 2: Delorme et al.

Figure 2. Overall reaction time distribution of go-responses in both the animal categorization task (gray traces and shaded distribution) and the recognition task (black lines). The top two traces are for correct go responses towards targets, the bottom two traces are for false alarms induced by non-target stimuli.
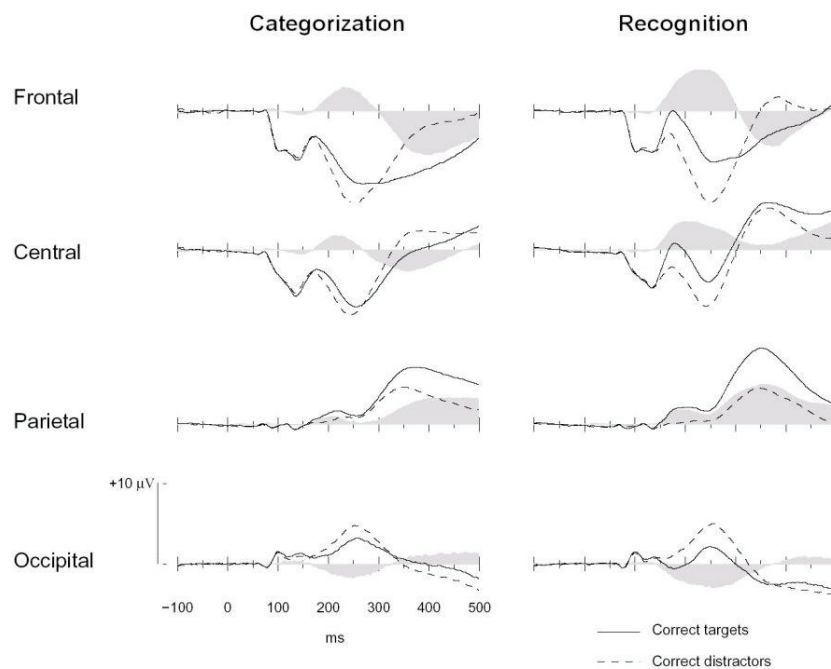

Figure 3: Delorme et al.

Figure 3. Grand average differential ERP activity. Average ERPs for all subjects in the categorization task (left column) and in the recognition task (right column) at different scalp locations: frontal, central, parietal and occipital sites corresponding respectively to the midline electrodes Fz, Cz, Pz and Oz. Average ERP on correct target trials (black line), average ERP on correct distractor trials (dashed lines), differential activity between correct target and distractor trials (shaded area). Note that the latency of the differential activity is always shorter in the recognition task.
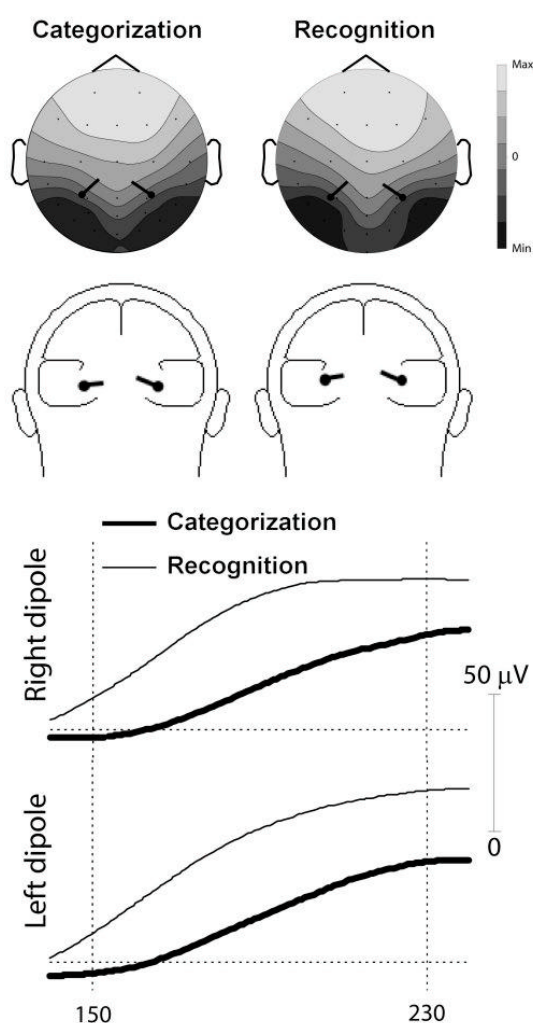
23

Figure 4. Cartography of the differential activity between the ERP waveforms of target and non-target data trials and localization of the electrical sources that accounted for this difference. For both tasks, the categorization task and the recognition task, a bilateral source accounted for more than 96 % of the differential ERP waveforms. Top: Gray-level scalp maps illustrate the averaged differential potential at 230 ms. Superimposed on these maps, the localization of the sources was virtually the same in both tasks. The location of the dipoles is also shown on frontal views. Bottom, the temporal dynamics of the left and right electrical source show that activation starts earlier and reaches a higher amplitude in the recognition task than in the categorization task.
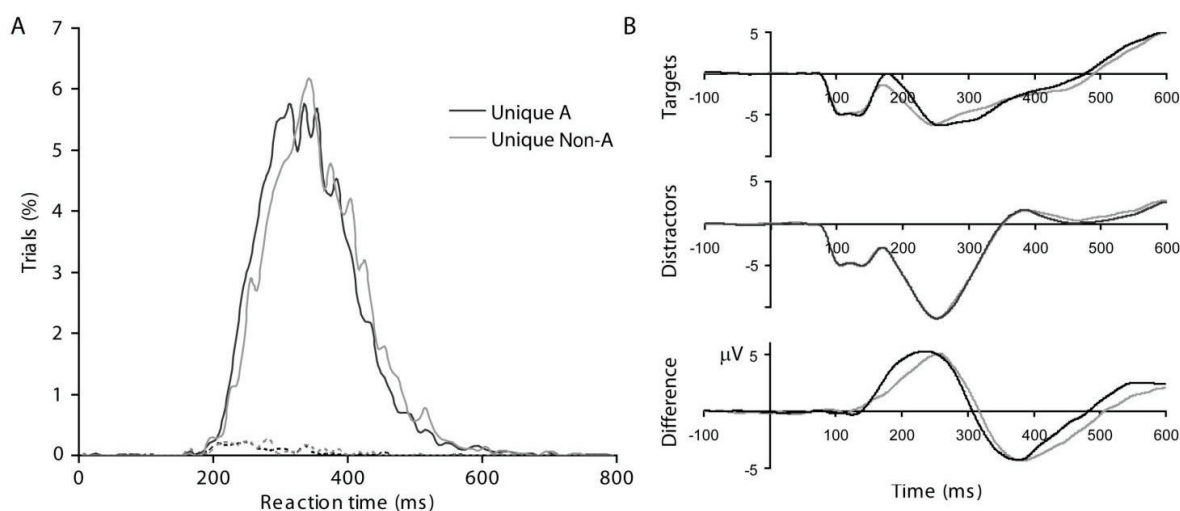
Figure 4: Delorme et al.

Figure 5: Delorme et al.

Figure 5. Overall results from the 14 subjects on the two different target-photograph sets in the recognition task. A, histogram of reaction time for the condition where pictures containing animals had to be recognized (Unique A) and for the condition where the target pictures did not contain animals (Unique Non-A). B, the differences between the frontal (FZ) ERPs recorded on correct targets trials (upper curves) and on non-targets trials (middle curves) are plotted (lower curves). In A and B: data are plotted in black for the animal set and in gray for the non-animal set.

24