Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 1459–1469
Vision Research
www.elsevier.com/locate/visres

The time course of visual processing: Backward masking and natural scene categorisation

Nadège Bacon-Macé *, Marc J.-M. Macé, Michèle Fabre-Thorpe, Simon J. Thorpe

Centre de Recherche Cerveau et Cognition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil, 133, Route de Narbonne, 31062 Toulouse, France

Received 23 April 2004; received in revised form 23 December 2004

Abstract

Human observers are very good at deciding whether briefly flashed novel images contain an animal and previous work has shown that the underlying visual processing can be performed in under 150 ms. Here we used a masking paradigm to determine how information accumulates over time during such high-level categorisation tasks. As the delay between test image and mask is increased, both behavioural accuracy and differential ERP amplitude rapidly increase to reach asymptotic levels around 40–60 ms. Such results imply that processing at each stage in the visual system is remarkably rapid, with information accumulating almost continuously following the onset of activation.

Keywords: Natural images; Backward masking; Early processing; Information integration; Event-related potentials (ERP)

1. Introduction

Human subjects are very quick and efficient at analysing briefly viewed natural scenes, an ability that has obvious survival value. We can determine whether a briefly flashed image contains an animal and make a behavioural response in as little as 250 ms, and this ability extends to other categories of visual stimulus such as faces or means of transport (Macé & Fabre-Thorpe, 2003; Rousselet, Macé, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a). Simultaneously recorded event-related potentials (ERP) diverge sharply between correct target and distractor trials just 150 ms after stimulus onset (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe et al., 1996) which imposed even more severe temporal constraints. Extensive training failed to reduce this 150 ms latency, indicating that even with images never seen before, the system is operating virtually optimally and with a minimal number of processing stages (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001).

This sort of behavioural and electrophysiological evidence imposes an upper limit on the amount of time required for animal detection but provides relatively little direct information about the dynamics of the underlying processing. With only a 150 ms delay between the onset of activation in the retina and a cerebral differentiation between target and distractor pictures, it is a challenge to explain how visual information is processed and transmitted through the visual pathways. A distinction is often made between discrete or continuous models of information transmission (Eriksen & Schultz, 1979; Hasbroucq, Burle, Bonnet, Possamai, & Vidal, 2002;

* Corresponding author. Tel.: +33 5 62 17 37 75; fax: +33 5 62 17 28 09.
E-mail address: nadege.bacon-mace@cerco.ups-tlse.fr (N. Bacon-Macé).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.01.004

McClelland, 1979), the first implying that there is a fixed minimum processing time at each stage before information can be sent to the next level, while the latter supposes that it can be transmitted continuously as soon as information becomes available. Both are consistent with a pipeline processing scheme in which every step can operate simultaneously and in parallel. Indeed, some form of pipeline processing seems necessary to account for the results of a recent study using RSVP (rapid serial visual presentation) showing that human subjects can detect images in sequences presented at rates of up to 75 images per second (Keysers & Perrett, 2002). Such data imply that less than 15 ms are enough to process a sufficient amount of information concerning each picture of the sequence.

RSVP experiments can be integrated into the broader approach of masking, which involves two or more temporally close stimuli to reduce the associated perception (Breitmeyer, 1984). Masking protocols are very useful to study the timing of information processing in the visual system since they allow processing to be interrupted at different times. Electrophysiological studies on monkeys have shown that the intensity and duration of neuronal responses are more and more affected as the mask gets closer to the stimulus, but that considerable information is available in monkeys from the first 30 ms of the neuronal responses (Rolls, Tovee, & Panzeri, 1999; Tovee & Rolls, 1995). In human subjects, there are many experiments that concern the influence of stimulus/mask interval on behavioural responses (Breitmeyer, 1984; Enns & Di Lollo, 2000), but few of them were used in the context of high-level tasks, such as categorisation. Moreover, few masking experiments have investigated the associated changes in cerebral activity, and most of those have involved fMRI methods. Nevertheless, there are reports of a correlation between the ability to detect or to name objects and the activation in occipital regions (Dehaene & Naccache, 2001; Grill-Spector, Kushnir, Hendler, & Malach, 2000; Vanni, Revonsuo, Saarinen, & Hari, 1996). As image and mask get temporally closer, both performance and cerebral activation decrease. This type of correlation can be particularly useful to understand the signals recorded from the brain during perceptual processing.

We present here the results of a backward masking experiment in a go/no-go categorisation task, in which natural scenes were followed by a very strong dynamic mask after a varying stimulus onset asynchrony (SOA). By interrupting processing after different delays, we could determine how information accumulates over time during the task. One of the novel features of the experiment was the use of a high screen refresh rate (160 Hz) that allowed us to present the test image for a single 6.25 ms frame and to vary the SOA by small 6.25 ms steps, a much higher resolution than is typically used in masking experiments.

2. Experimental procedure

2.1. Task

Sixteen subjects participated in this experiment (8 females, 8 males, with a mean age of 29 ranging from 21 to 50). They all volunteered for the study and gave their written informed consent. The go/no-go categorisation task was based on an experimental procedure introduced by Thorpe et al. (1996). Subjects were seated in a dimly lit room, at 1 m from a screen adjusted to an 800×600 pixel resolution and a 160 Hz refresh rate. Natural scene pictures (600×400 pixels in size) were flashed on the monitor for a single frame, which corresponds to 6.25 ms. Subjects were asked to release a button within 1 s if the picture contained an animal and maintain pressure otherwise.

Each subject was tested on 16 series of 90 trials, each of which contained the same number of target and distractor images. All subjects had previously completed at least 3 training blocks of 90 trials. They were asked to try to release the button on 50% of the trials, whatever the masking condition.

A trial began with the display of a white fixation cross in the middle of the black screen for 600–900 ms at random. Then the picture—target or distractor—was flashed, followed by the mask stimulus. Eight different values were used for the stimulus onset asynchrony (SOA) between the picture and the mask (6.25, 12.50, 18.75, 25.00, 31.25, 43.75, 81.25 and 106.25 ms) and display latencies were verified with a photodiode connected to an oscilloscope. Furthermore, we added a control condition in which only the mask was displayed after the fixation cross, without any picture presentation. These 9 conditions were counter-balanced in each series, with 10 trials per condition presented at random, producing a total of 90 trials per block. Any given subject only saw each picture once.

χ^2 tests were used to evaluate if behavioural accuracy was above chance level for each SOA condition. Masking effects between the conditions were assessed with analysis of variance (ANOVA) and post-hoc analyses were performed by using paired t-tests with a Bonferroni correction or Mann–Whitney tests.

2.2. Stimuli

A total of 1280 grey level natural images were used in this experiment. As demonstrated in previous work, ultra-rapid categorisation does not rely on colour cues, as performance is almost unaltered when stimuli are presented in grey level (Delorme, Richard, & Fabre-Thorpe, 2000). Moreover, masking effects were easier to obtain and control without colour information in the natural scenes. Half the images contained animals, and were as varied as possible (fish, insects, mammals,

birds or reptiles). The subjects had no knowledge about the size, position and number of the targets in a single picture. The other half of the images were distractors with a wide range of material including natural landscapes, indoor or outdoor scenes, man-made objects, etc... None of the pictures had been seen previously by the subjects and training pictures were not used in the test series.

2.3. Mask

To construct the mask, a white noise image was filtered at four different spatial scales, and the resulting images were thresholded to generate high contrast binary patterns. For each of the 4 spatial scales, 4 different versions were generated by mirroring and rotating the original image. A pool of 16 images was thus available for masking. The mask used in this experiment was a sequence of 8 images - a so-called "dynamic mask" (Fig. 1). The 8 images were chosen randomly from the pool, with each of the four spatial scales presented once during the first 4 images and again during the last four images. Thus, a pattern at each of the spatial scales appeared twice in the "dynamic" mask (see Fig. 1). All the images in the mask were presented for 2 refresh cycles, so that overall, the masking stimuli were displayed for 16 frames (around 100 ms).

2.4. ERP recordings

EEG data were recorded from a 32-electrode cap. Electrode locations were defined using the standard 10–20 Oxford system with 12 additional electrodes over occipital sites. Electrical activity was amplified by a NeuroScan Synamps amplifier linked to a PC computer, digitized at 1000 Hz, corresponding to a sample bin of 1 ms, and low-pass filtered at 100 Hz. Each recording epoch began 100 ms before the stimulus display on the screen and continued for 1000 ms after the stimulus onset. A baseline correction was carried out for each epoch using the 100 ms of pre-stimulus activity. Trials with artefacts related to ocular movements were rejected, by using a criterion of $[-80;+80\ \mu V]$ on two frontal electrodes (FP1 and FP2) between −100 and +400 ms. Within this time period, another artefact rejection was performed on trials with a strong alpha frequency activity, by using a $[-40;+40\ \mu V]$ criterion on parietal electrodes (Oz and Pz). Signals were then low-pass filtered at 40 Hz before the analysis. We were particularly interested in the occipito-temporal electrodes (standard O1, O2, OZ, IZ and non-standard PO7, PO8, PO9, PO10, O9, O10, P7, P8) and the frontal electrodes (standard FP1, FP2, F3, F4, Fz). Epochs corresponding to correct responses were averaged separately for targets and distractor trials on each masking condition.

Differential activities were determined by subtracting the average signal on correct distractor trials from the signal on correct target trials. Eight different curves were obtained, one for each SOA condition. The differential activity amplitude was calculated by a Matlab program that determined the most negative point between 150 and 250 ms after the onset of stimulus presentation (Rousselet, Thorpe, & Fabre-Thorpe, 2004; Thorpe et al., 1996). It was measured separately for each individual and also using the average signal across all

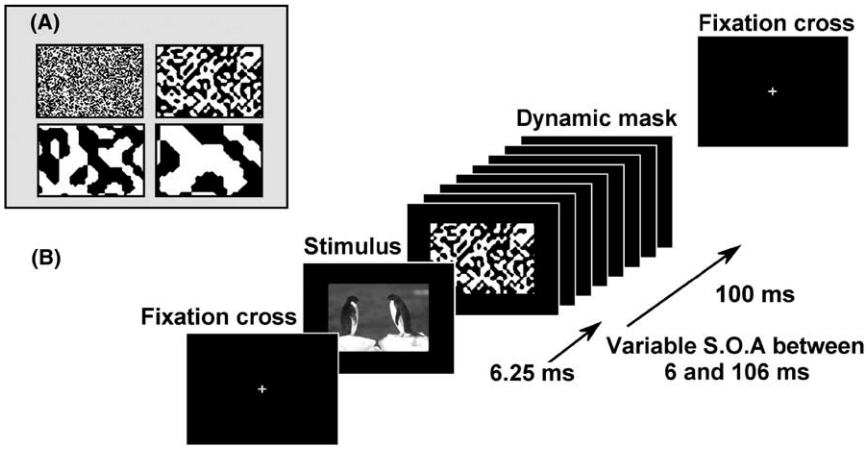

Fig. 1. Behavioural paradigm. (A) Four pictures with different spatial scales that constitute the dynamic mask. Each could be presented at 4 different orientations making a total of 16 different patterns. The images were intermixed to reduce the risk of generating retinal after-effects with the restriction that each spatial scale was used once during the first 4 pictures of the mask and once again during the 4 last ones. (B) In each series, subjects were tested on 90 trials organised as follows: first the fixation point is displayed on the centre of the screen for a random delay to avoid anticipated responses. Then the grayscale picture is flashed for only one frame using a monitor set at 160 Hz. After a variable 6.25–106.25 ms SOA, a dynamic mask is displayed, composed of eight 100% contrasted mask patterns at the four different spatial scales. The subjects then have 1000 ms to release the button if the picture contains an animal. Eight time steps were chosen for the SOA: 6.25, 12.50, 18.75, 25, 31.25, 43.75, 81.25 and 106.25 ms.

subjects. Differences in latencies and amplitude among SOA conditions were statistically evaluated by ANOVAs and post-hoc analyses were performed using *t*-tests with a Bonferroni correction. The correlations between electrophysiological measurements and behavioural data were performed with Pearson tests.

3. Results

3.1. Behavioural performance

3.1.1. Mask efficiency

We evaluated behavioural performance in terms of accuracy and reaction time as a function of the SOA. For each condition, a χ^2 test between correct and incorrect responses determined if accuracy was above chance level, set at 50% because targets and distractors were equally likely. Only the very shortest SOA interval (6 ms) resulted in performance at chance level, with a mean value of 51.9% (Fig. 2A). This result emphasizes the high efficiency of the mask, which effectively prevented visual processing when presented close to the stimulus. However for the next SOA (12 ms) condition, accuracy was already above chance level ($p < 0.01$) and rapidly increased to reach 85.6% with a 44 ms SOA. Accuracy then stabilized at a maximum value of 91.4% for the last condition (106 ms). However, increasing processing time above 44 ms had only a minor effect on performance since accuracy scores in the last three SOA conditions 44–81–106 ms were not significantly different ($p > 0.33$). Note that the maximum accuracy was very close to the accuracy obtained in a previous study (Delorme et al., 2000), using achromatic natural images flashed for 20 ms in the same go/no-go categorisation task without masking, and where subjects averaged 93% correct. Thus the mask has relatively little effect when it appears 40–60 ms after the image presentation onset, and visual processing remains extremely good despite the fact that the stimulus picture was flashed for only 6 ms.

3.1.2. Response inhibition with increasing difficulty

We noticed a strong reduction in response rate with the most difficult masking conditions. Before the experiment, subjects were asked to try to release the button on about half of the trials in each series. This instruction has been respected since the mean response rate, including correct and incorrect responses, was about 47% when grouped across all conditions. However, the response rate varied considerably with the SOA, as it exceeds 50% from 106 to 25 ms SOA, and drops strongly with SOAs below 25 ms (Fig. 2B).

Interestingly, these variations only affected the proportion of correct go responses. In contrast, the proportion of erroneous go responses to distractors was remarkably stable across all SOA conditions. It would appear that short SOAs did not lead subjects to make more false positives to distractor pictures but did prevent them extracting enough information to make a response on target trials. Another interesting result is given by comparing the response rate obtained with the shortest SOA (22.3%) with the control condition when only the mask was displayed without any picture (19.6%). These two conditions were not significantly different ($p > 0.24$) which suggests that with a 6 ms SOA, subjects behaved as if no image had been presented at all.

3.1.3. Reaction times

Mean reaction time decreased with longer SOAs, particularly for values over 44 ms ($F(7,127) = 2.591$, $p < 0.02$). In the conditions where the mask was close to the stimulus (SOA 6, 12, 19, 25 and 31 ms), reaction times were significantly longer ($p < 0.01$) than when the mask appeared later (44, 81 and 106 ms). A maximum difference of 54 ms was found between the 12 and 81 ms SOA conditions. This suggests that when the mask interrupts the visual processing earlier, the amount of

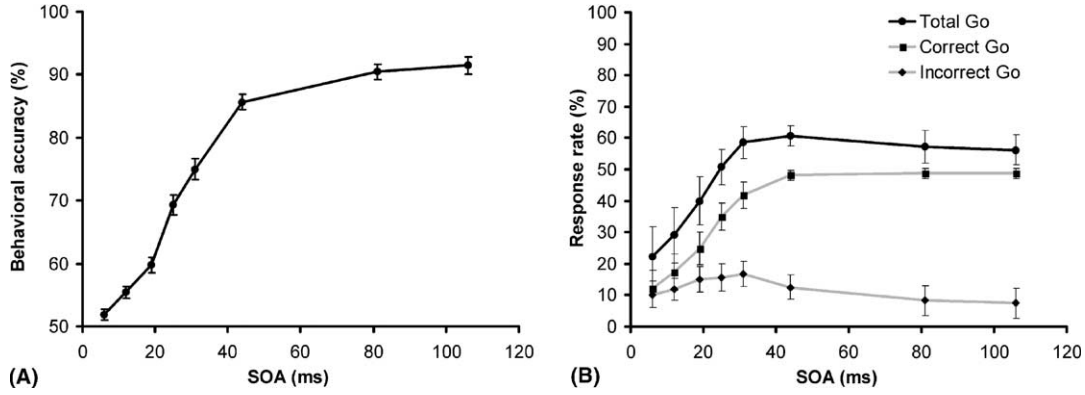

Fig. 2. Behavioural performance as a function of the SOA, averaged above 16 subjects. (A) Behavioural accuracy (±s.e.m.). (B) Mean percentage of go responses (±SD).

information is reduced and perceptual decisions require more time.

Masking effects can be observed throughout reaction time distributions by comparing the condition of optimal perception (106 ms) with each of the others (Fig. 3A). Above 44 ms, distributions are remarkably similar, but with 31 ms SOA, the median part of the distribution shows a pronounced plateau. Thus, there is a strong effect of the mask on the median reaction time, but the initial part of the distribution, corresponding to early responses, is not affected. With an SOA of 25 ms and less, early responses are also disrupted and the early part of the reaction time distributions no longer superimpose.

3.2. Electrophysiology

Disruptive effects on visual processing can also be observed on the ERP data. For each subject we averaged separately signals on distractor and target trials and subtracted one from the other to calculate a differential activity curve. As signals on targets and distractors contain information about the response to both the picture and the mask, subtracting these two signals is a good way to cancel out the activity associated with the physical encoding of the mask. The effects of the mask on image categorisation processing remain clearly observable on the residual signal (Fig. 4). Therefore, we will not present here a detailed analysis of the shape of the underlying ERP signals but rather focus on an analysis of the differential effects.

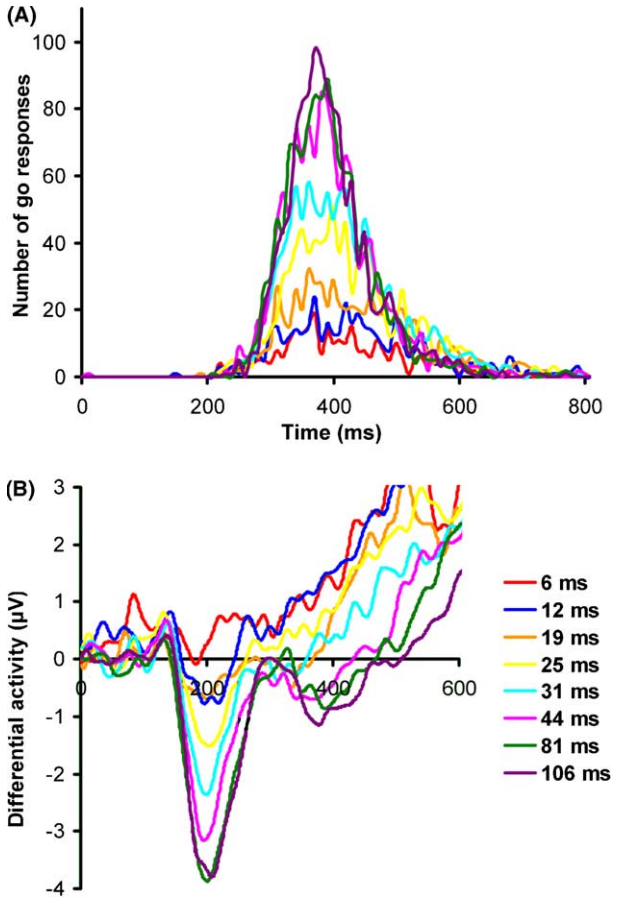

Fig. 3. Masking effects on behavioural reaction time and occipital cerebral activity. (A) Reaction time distribution of correct go responses as a function of the SOA (10 ms bin width), averaged on 16 subjects. (B) Differential activity averaged on 16 subjects for each SOA. The activity is calculated as the difference between signals on correct target and distractor trials, obtained from PO8 electrode.

We analyzed the differential activity with respect to the SOA. Fig. 3B shows averaged signals recorded on a representative occipital electrode (PO8). The onset of the differential activity appears to start at around the same latency (150 ms) but it is clear that the signal amplitude decreases with shorter SOAs ($F(7,1535) = 77.13$, $p < 0.001$). Moreover, with the exception of the two shortest SOAs (6 and 12 ms) for which the activity was rather weak, peak latencies are remarkably stable between 200 and 215 ms. In other words, when the mask is closer and closer to the picture, the reduction of perceptual differences between target and distractor stimuli strongly affects the amplitude of differential activity.

A lateralization effect can be observed in this task, as the mean amplitude of the differential activity was significantly larger for the electrodes over the right hemisphere compared to the left hemisphere (respectively 2.56 μV for the average of electrodes O2, PO10, PO8, O10, P8 versus 2.22 μV for the average of O1, PO9, PO7, O9, P7, all SOA conditions grouped; $F(1,1279) = 15.45$, $p < 0.0001$). This was the case for each of the SOA conditions except for the shortest one at 6 ms.

Although the masking effect is particularly visible on occipital electrodes, similar effects can be seen at most electrode sites (Fig. 5). At frontal sites, shortening the SOA induced a significant diminution of the maximal amplitude of the differential activity ($F(7,639) = 22.305$, $p < 0.0001$), appearing around 200 ms. But in contrast no lateralization effect could be observed at these sites ($p = 0.145$).

These electrophysiological results can be directly related to behavioural data, which also showed a clear diminution of performance with decreasing SOA. Differential activity amplitude and behavioural accuracy variations are in fact strongly correlated, as showed on Fig. 6. This observation reinforces the idea that the differential activity reflects the result of a perceptual decision and that differential ERP responses provide a powerful investigative tool.

3.3. Control experiment

The choice of the dynamic mask was made after a number of pilot experiments, and appeared to be very

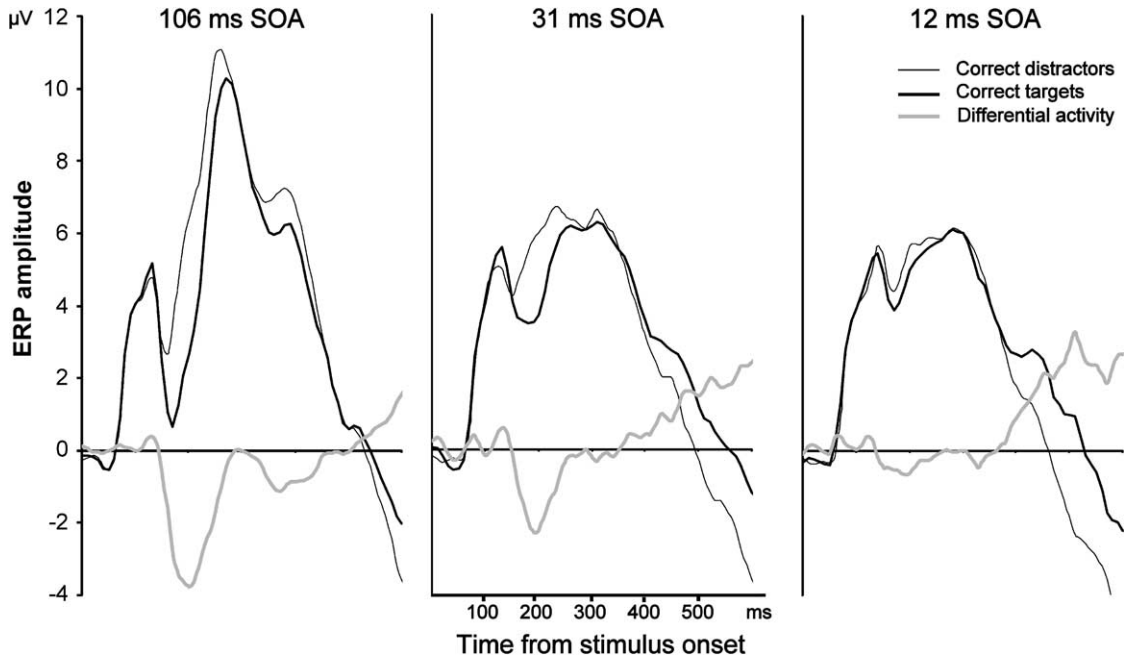

Fig. 4. Grand-average ERP on electrode PO8 for three SOA conditions. Differential activity is obtained by subtracting the signals on correct target and distractor responses.

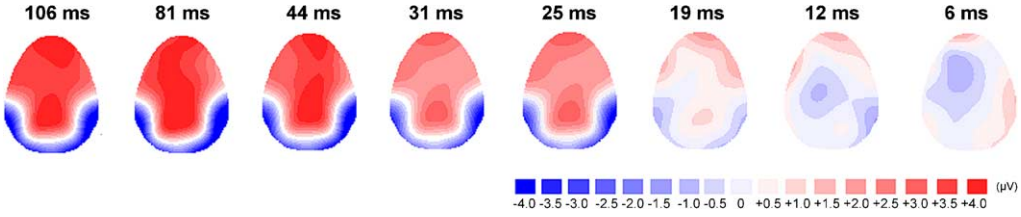

Fig. 5. Differential activity over the scalp at the maximal point of the amplitude, 200 ms after the image onset. Activity was averaged over 16 subjects.

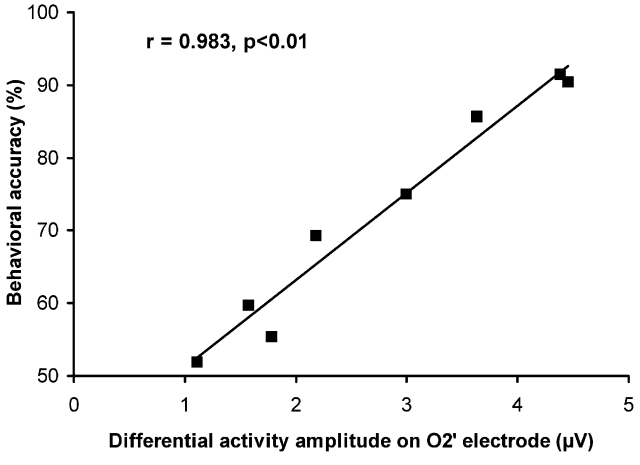

Fig. 6. There is a linear relation between behavioural accuracy and the amplitude of the occipital differential activity, particularly on the right hemisphere. Differential activity from PO8 electrode has been averaged over the 16 subjects, and correlated by a Pearson test ($p < 0.01$) to the mean behavioural accuracy among all SOA conditions.

efficient, as demonstrated by the fact that performance was at chance level with the shortest SOA. However, one possible problem of using a mask with multiple-frames is that one cannot be sure at what exact point the masking becomes totally effective. This could potentially add considerable uncertainty to the SOA considered as the disruptive latency.

We therefore have made a behavioural control experiment where 10 subjects performed the categorisation task at four different SOA values (4 conditions: 6, 12, 44 and 106 ms) and in which we varied the number of pictures in the mask from 1 to 8 (5 conditions: 1, 2, 3, 4 and 8 pictures). Across all the subjects, a total of 640 trials was performed for each of the 20 conditions. Furthermore, we encoded the spatial scale pattern used for each picture of the mask and particularly for the first one. These patterns were randomly chosen for each trial. Fig. 7 shows behavioural accuracy as a function of the first mask pattern. With only one image in the mask

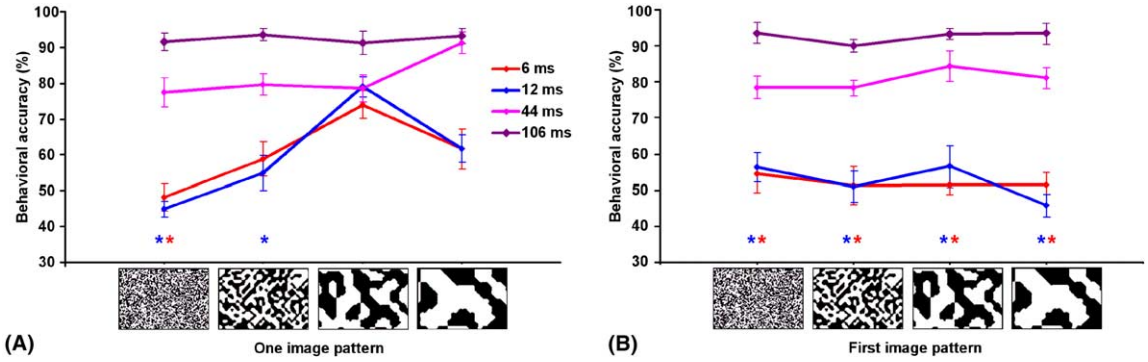

Fig. 7. Behavioural accuracy as a function of the spatial scale of the mask, for different SOAs (±s.e.m.) and as a function of the number of images in the mask. (A) Results when only one image is used in the mask. (B) Results when a sequence of two images is used in the mask. The pattern of the first image is illustrated here, all 3 other spatial patterns were used equally as the second image of the mask. The different curves represent behavioural accuracy for different SOAs. Ten subjects performed the experiment. Coloured asterisks indicate when the condition is at chance-level ($p < 0.01$).

(Fig. 7A), disruption effects depended strongly on the spatial scale of the mask at 6 and 12 ms SOA. When the finest spatial scale was used, masking was effectively complete since performance was at chance-level. Even with the second finest scale, performance was still very poor. This means that on virtually half the trials, just the first mask pattern was enough to completely disrupt processing, limiting the number of problematic trials. In contrast, at coarser spatial scales, the masking was less effective and subjects were able to perform more than 70% correct when just the mid-to-coarse scale mask image was used. However, when the mask contained two different images in succession (Fig. 7B), accuracy was no better than chance for both 6 and 12 ms SOA, irrespective of which spatial scale was flashed first. This result may appear to contradict the data from the previous experiment, in which performance was significantly above chance with an SOA of 12 ms, despite having used an even longer 8 image dynamic mask. However, it should be noted that there were much more trials per condition in the original experiment (2560 vs. 640 trials), which increases the statistical sensitivity of the test.

Together, the results of this control experiment demonstrate conclusively that the masking effects were indeed very strong and occurred very rapidly from the onset of the mask. Given that for the main experiment, the 8 image dynamic mask was continued for 100 ms, we can safely conclude that the disruption was complete throughout a critical period for target processing.

4. Discussion

4.1. Time course of information extraction

4.1.1. Visual information is extracted before masking

As might be expected, the behavioural data shows a strong masking effect on the visual processing involved in this high-level categorisation task, with both a drop in accuracy and an increase in reaction times. With a 6 ms SOA, processing appears to be completely blocked since the subjects were unable to perform significantly above chance. However, from 12 ms onwards, accuracy is already above chance level and performance increases until a ceiling effect is reached between 44 and 81 ms.

The high vertical refresh rate of the monitor (160 Hz), allowed us to observe a large range of masking levels, which leads to make several remarks about visual information extraction. First, it is noteworthy that the maximum accuracy reached by the subjects was close to the accuracy obtained in the same categorisation task used without masking (Delorme et al., 2000), indicating that the mask has no major effect after 81 ms. This also means that the presentation time of the picture, reduced to 6 ms in the present experiment from 20 ms in most of our previous experiments, had no appreciable effect on precision. It therefore appears that this 6 ms stimulation period is sufficient for the retina to extract enough information from the picture for animal detection to occur. Finally, the control experiment confirmed that the masking effects were strong even with just the first mask image, and that by the time the second mask pattern was presented, the disruption was complete. Given that the dynamic mask was maintained for 100 ms, we can be very confident that a long period of target processing is affected by the effective masking. This supposes that the delay available to extract relevant features from the stimulus is limited before masking takes place, and the visual system should base its analysis on a restricted amount of information to perform the task.

4.1.2. Visual information accumulates over time

The electrophysiological data also argue for a progressive accumulation of information. The differential activity, calculated by taking the difference between ERPs on correct target and distractor trials, is strongly affected by the SOA reduction. Its amplitude decreases

when the mask gets closer to the picture. We found a very high correlation between this reduction and behavioural accuracy. This can be related to another analysis of the differential activity in a previous categorisation task (Rousselet et al., 2004), where the status of the trials (Correct / False Alarm / Missed) could also be linked to the amplitude of the differential effects at occipital, frontal and parietal sites. The predictability of the behavioural outcome on the basis of the differential ERP signals constitutes a striking demonstration of a strong link between perception and cerebral activity, as previously shown with both fMRI (Dehaene & Naccache, 2001; Grill-Spector et al., 2000; Vanni et al., 1996) and unitary recordings in monkeys (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Leopold & Logothetis, 1996; Thompson & Schall, 1999).

The analysis of the differential signals between roughly 150 and 250 ms demonstrated that the more the activity averaged on targets differs from the distractor activity, the more subjects are able to detect the animals. This result could be related to the analysis of response rate since the mask has a higher effect on subjects' decisions for targets than for distractors (Fig. 2B). The overall data show that the difference between target and distractor responses is maximized when the mask is presented far from the stimulus, as if there were more and more cues accumulating to dissociate these two groups of images. The results are in accordance with the model of sensory information accumulation proposed by Schall (Schall, 2001) and derived from earlier work by Shadlen, Newsome and colleagues (Gold & Shadlen, 2000; Kim & Shadlen, 1999; Salzman & Newsome, 1994; Shadlen & Newsome, 1996). In their experiments, monkeys were trained to judge the main direction of motion in a collection of moving dots. The monkeys reported their responses by making an eye movement to one of two points, each indicating a given direction. The authors proposed that the decision depends on the accumulated signal corresponding to the increasing discrimination of the main direction of motion. In the same way in our experiment, we can suppose that the greater the separation between stimulus and mask, the greater the amount of processing that can be performed. Relevant information concerning the presence of an animal in the picture is accumulated until reaching a decisional threshold.

This model of cue accumulation over time fits with our data on reaction time. With SOAs between 25 and 44 ms, the early part of the reaction time distribution did not appear to be affected by the mask, but we observed a saturation effect on correct responses with median reaction times (Fig. 3A). This suggests that when pictures contain particularly salient cues, extensive information accumulation is not necessary and the subjects are capable of executing fast responses. However, when the target discrimination requires more analysis, information would not be available because processing is disrupted by the mask. Below 25 ms, the integration time was probably insufficient to process as much information because the mask affected even the earliest responses.

If the visual system bases the results of its analysis on the accumulated information, what does it imply for information encoding? How can the mechanisms of information extraction at different steps of the visual pathways be decomposed and what determines the impact of masking interference?

4.2. Information encoding in the visual pathways

4.2.1. Interference between stimulus and mask information: the where and how issues

Behavioural performance does not increase much with SOAs longer than 40 ms. We may relate this result with the latencies obtained from macaque neurophysiology showing that the first 30–40 ms includes the most selective part of the neuronal responses (Kovacs, Vogels, & Orban, 1995; Rolls et al., 1999; Tovee & Rolls, 1995). This data suggests that there is an upper limit on the time required at each processing stage to extract the relevant information that needs to be transmitted to the next step. Any processing that would take longer would be obliterated or smothered by the mask information.

Where would these masking effects take place? A first model would propose that the effects are more likely to occur relatively early in the visual pathways, for instance at the level of V1, depending directly on the structure where mask information could be encoded. Recordings in monkey infero-temporal cortex have demonstrated that the majority of neurons are maximally activated by stimuli more complex than bars or simple textures (Tanaka, Saito, Fukada, & Moriya, 1991), and showed a high degree of sensitivity to image scrambling, with activation decreasing together with performance (Vogels, 1999a, 1999b). Other studies, using functional imaging in humans, have compared the activation produced by objects and textures and found that a region of lateral occipital cortex was preferentially activated by objects even when the spatial frequencies and contrast of the object stimuli matched those of the texture stimuli (Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998; Malach et al., 1995). All these studies suggest that the mask, as a kind of texture stimulus, should mainly interact with picture information in earlier areas.

The next question concerns the *mechanism* by which the masking effect occurs. One simple explanation of masking assumes that the mask produces interference when neural responses to the mask and the test image overlap in time, and this effect is all the more important when it concerns spatially overlapping information related to critical features of the stimulus. There is good

evidence that the activation of sensory inputs to areas such as the striate cortex results in strong intracortical inhibition that could well interfere with the processing of subsequent inputs. The disrupting effects will thus depend on the spatio-temporal overlap between the neural responses to the test and mask stimuli.

Neurophysiological studies have shown that the onset latencies of neurons within a given visual structure vary from neuron to neuron, even when the visual stimulus is unchanged. For example, in primate visual cortex, onset latencies can vary from as little as 30 ms to 70 ms or more. The reasons for this variability are diverse, but one of the most important factors is undoubtedly stimulus contrast. It is notable that the shortest latencies ever seen have been obtained with very high contrast and high luminance stimuli. Given that the mask stimuli used in our experiments all have maximal contrast, we can assume that many neurons in V1 will respond to the mask with particularly short latencies (Albrecht, Geisler, Frazor, & Crane, 2002; Albrecht & Hamilton, 1982; Foxe & Simpson, 2002; Nowak & Bullier, 1997; Reich, Mechler, & Victor, 2001; Sestokas & Lehmkuhle, 1986). In contrast, the neural responses to the natural images used as test patterns are likely to be substantially more variable. Indeed, if we suppose that any given photograph of an animal will contain many different features that can be diagnostic for the presence of an animal, it is clear that the contrast associated with each feature will vary a lot. Thus, much of the critical information about the stimulus will be conveyed less rapidly than information about the mask, strengthening the effects of inhibitory mechanisms. Only information that can survive this spatio-temporal overlap would then be transmitted to the next step, and contribute to accumulate cues about the test stimulus. The first interpretation is thus based on the disruption of feed-forward processes, due to mask processing catching up with stimulus processing.

Another interpretation is based on the difference in transmission rates along the fast magnocellular (M) and the slower parvocellular (P) visual pathways. Detailed chromatic representation in the P stream reaches visual cortex roughly 20 ms after the M inputs that mainly transmit motion and coarse luminance-based information (Nowak & Bullier, 1997; Nowak, Munk, Girard, & Bullier, 1995). Taking into account this 20 ms delay between the two streams of information, the mask might have little effect on the magnocellular processing of the test image but would strongly interfere with its parvocellular processing. However, magnocellular information may be sufficient to allow good accuracy in the fast categorisation task used here (Delorme et al., 2000; Macé, Thorpe, & Fabre-Thorpe, in press), and the interference of the highly contrasted mask with the feed-forward processing of magnocellular information (Macé et al., in press) may still be significant.

Finally, mask processing could interrupt feedback processing of the stimulus, at least at two levels. Iterative loops are thought to be important for segmentation, and involve the convergence of feedback from higher areas to areas like V1 or V2 (Hupe et al., 1998; Lamme, Super, & Spekreijse, 1998). In such pattern masking experiments, the processing of feedback information is probably made difficult with a mask closely following the stimulus. Moreover, subjects often reported that they released the button without explicit understanding of the photograph, which is in accordance with the common idea that feedback processing may be crucial for conscious image perception (Bullier, 2001; Lamme & Roelfsema, 2000; Pascual-Leone & Walsh, 2001).

These interpretations are not mutually exclusive and could even explain the striking differences between the effect of short SOAs on the initial part of the RT distribution and the plateau effect obtained with the 31 ms SOA (Fig. 3A). In fact, two different kinds of perturbations may be reflected here. The plateau effect may result from disruption of feedback processing related to the detection of the target, or disruption of direct parvocellular inputs. In contrast, the shift observed in the initial part of the RT distribution with shorter SOAs could reflect the disruption of the initial wave of processing.

4.2.2. A pipeline architecture

If we assume that the visual system accumulates sensory information until a decision threshold is reached, the very progressive masking effect over time is another point of interest. Presumably, this sort of task requires information processing at several different levels of the visual system including the retina, LGN, V1, V2, V4 and inferotemporal cortex. We have argued in the past that this sort of fast processing may leave only a short time at each processing stage before the next level has to respond, maybe as little as 10 ms or so (Bullier & Nowak, 1995; Fabre-Thorpe et al., 2001; Nowak et al., 1995; Thorpe & Fabre-Thorpe, 2002). These results confirm that visual processing can rely on such short latencies and challenges traditional views that use firing rate codes to convey information (Thorpe, Delorme, & Van-Rullen, 2001; VanRullen & Thorpe, 2001b; VanRullen & Thorpe, 2002). Further investigations will be necessary to understand how visual processing can be performed in such temporally constrained conditions. In the case of a serial model of information transmission, we might have expected sharply contrasted responses in which performance and ERPs are strongly affected below the decision threshold and less disrupted above this threshold, although the averages of performance across trials and subjects may obscure some types of more discrete transition (Miller, 1988). In contrast, in the case of a continuous model, masking effects may be progressively lessened with increased SOA. Our data showed that as SOA is increased above 12 ms, subjects

gradually increased their ability to detect an animal in the picture, which suggests that information can be conveyed to extrastriate areas in a continuous and asynchronous way. The lack of a clear threshold may also indicate that the information does not need to be fully processed at each stage (that is for all points of the space at the same time), but could be forwarded to the next stage and processed more progressively. Although the results do not argue conclusively in favour of a continuous model, they nevertheless strongly suggest that information transfer occurs progressively at each stage using a form of pipeline architecture.

A final point concerns the nature of the information used to perform the animal/non-animal task. While, in principle, we think that this should be considered as a true high-level visual task, there have been suggestions that even relatively high-level categorisations such as "natural vs man-made scenes" can be made on the basis of relatively low-level information. For example, Torralba and Oliva have reported that a linear combination of the outputs of a series of orientation and spatial frequency tuned channels can allow performance at over 80% correct (Torralba & Oliva, 2003). We certainly cannot exclude the possibility that our subjects are using this sort of information. However, so far at least, none of these purely low-level strategies has succeeded in achieving performance levels of above 90%, nor have they been used to differentiate between classes of objects. We therefore feel that other more complex visual processing strategies are probably at work. The current set of experiments does not allow us to distinguish between these possibilities. Nevertheless, they do demonstrate that, whatever the nature of the information used to perform the task, it is information that the visual system can extract extremely rapidly.

Acknowledgements

This work was supported by the CNRS, the ACI Integrative and Computational Neuroscience. Financial support was provided to both N. Bacon-Macé and M.J-M. Macé by a Ph.D. grant from the French government. We thank G.A. Rousselet for technical assistance and helpful discussions on the results of the experiment.

Supplementary data

Correlation between behavioural accuracy and the amplitude of the occipital differential activity, on the right hemisphere occipital electrodes. The values shows individual correlation calculated with a Pearson test ($p < 0.01$) between behavioural accuracy and the differential activity amplitude on five occipital electrodes. Differential activity amplitude was determined by the most negative point between 150 ms and 250 ms on averaged signals by condition, for each subject. The bottom line indicates an even stronger r-value by averaging the parameters for all 16 subjects before correlating them. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.visres.2005.01.004.

References

Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *Journal of Neurophysiology, 88*(2), 888–913.

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology, 48*(1), 217–237.

Breitmeyer, B. G. (1984). *Visual masking: an integrative approach* (p. 454). Oxford, New York: Oxford University Press.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience, 13*(1), 87–100.

Bullier, J. (2001). Feedback connections and conscious vision. *Trends in Cognitive Sciences, 5*(9), 369–370.

Bullier, J., & Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology, 5*(4), 497–503.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*(1–2), 1–37.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research, 40*(16), 2187–2200.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*(9), 345–352.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Perception and Psychophysics, 25*(4), 249–263.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*(2), 171–180.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. *Experimental Brain Research, 142*(1), 139–150.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404*(6776), 390–394.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron, 21*(1), 191–202.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*(8), 837–843.

Hasbroucq, T., Burle, B., Bonnet, M., Possamai, C. A., & Vidal, F. (2002). Dynamique du traitement de l'information sensorimotrice: apport de l'electrophysiologie. *Canadian Journal of Experimental Psychology, 56*(2), 75–97.

Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature, 394*(6695), 784–787.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*(3), 120–125.

Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience, 2*(2), 176–185.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of National Academic Science USA, 92*(12), 5587–5591.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience, 23*(11), 571–579.

Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*(4), 529–535.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature, 379*(6565), 549–553.

Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (in press). Rapid categorisation of achromatic natural scenes: how robust at very low contrasts? *European Journal of Neuroscience*.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of National Academic Science USA, 92*(18), 8135–8139.

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. *Psychological Review, 86*(4), 287–330.

Miller, J. (1988). Discrete and continuous models of human information processing: theoretical distinctions and empirical results. *Acta Psychologica, 67*(3), 191–257.

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. S. Rockland, J. H. Kaas, & A. Peters (Eds.). *Extrastriate visual cortex in primates* (Vol. 12, pp. 205–241). New York: Plenum Press.

Nowak, L. G., Munk, M. H., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neurosciences, 12*(2), 371–384.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510–512.

Reich, D. S., Mechler, F., & Victor, J. D. (2001). Temporal coding of contrast in primary visual cortex: when, what, and why. *Journal of Neurophysiology, 85*(3), 1039–1050.

Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: information analysis. *Journal of Cognitive Neuroscience, 11*(3), 300–311.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*(7), 629–630.

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*(6), 440–455.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Research, 44*(9), 877–894.

Salzman, C. D., & Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science, 264*(5156), 231–237.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *National Review of Neuroscience, 2*(1), 33–42.

Sestokas, A. K., & Lehmkuhle, S. (1986). Visual response latency of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. *Vision Research, 26*(7), 1041–1054.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of National Academic Science USA, 93*(2), 628–633.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66*(1), 170–189.

Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during masking. *Nature Neuroscience, 2*(3), 283–288.

Thorpe, S., Delorme, A., & Van-Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Network, 14*(6–7), 715–725.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Thorpe, S. J., & Fabre-Thorpe, M. (2002). Fast visual processing and its implications. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed.). Cambridge, MA: MIT Press.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14*(3), 391–412.

Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition, 2*(1), 35–58.

Vanni, S., Revonsuo, A., Saarinen, J., & Hari, R. (1996). Visual awareness of objects correlates with activity of right occipital cortex. *Neuroreport, 8*(1), 183–186.

VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird. Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30*(6), 655–668.

VanRullen, R., & Thorpe, S. J. (2001b). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation, 13*(6), 1255–1283.

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42*(23), 2593–2615.

Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part1: behavioural study. *European Journal of Neuroscience, 11*(4), 1223–1238.

Vogels, R. (1999b). Effect of image scrambling on inferior temporal cortical responses. *Neuroreport, 10*(9), 1811–1816.