Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 46 (2006) 1762–1776

Vision Research
www.elsevier.com/locate/visres

Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited ☆

Holle Kirchner *, Simon J. Thorpe

*Centre de Recherche Cerveau et Cognition (UMR 5549), CNRS - Université Paul Sabatier Toulouse 3, Faculté de Médecine de Rangueil,
31062 Toulouse Cedex 9, France*

Received 16 May 2005; received in revised form 28 September 2005

Abstract

Previous ultra-rapid go/no-go categorization studies with manual responses have demonstrated the remarkable speed and efficiency with which humans process natural scenes. Using a forced-choice saccade task we show here that when two scenes are simultaneously flashed in the left and right hemifields, human participants can reliably make saccades to the side containing an animal in as little as 120 ms. Low level differences between target and distractor images were unable to account for these exceptionally fast responses. The results suggest a very fast and unexpected route linking visual processing in the ventral stream with the programming of saccadic eye movements.

Keywords: Early visual processing; Eye movements; Natural images; Visual pathways; Event-related potentials

1. Introduction

The human visual system is very fast and efficient at extracting information about the objects present in complex natural scenes (Potter, 1976; Thorpe, Fize, & Marlot, 1996). The speed of this visual processing can be assessed both by behavioural and neuromagnetic measurements (Thorpe & Fabre-Thorpe, 2002; Thorpe et al., 1996). In go/no-go categorization tasks, human participants initiate manual responses with average and minimum reaction times of about 450 and 250 ms, respectively. However, these measures includes both the time of visual processing and response execution (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe et al., 1996). One approach to estimate the time for just the visual processing component alone involves analyzing simultaneously recorded event-related potentials that show differential effects starting to diverge at around 150 ms, sometimes even substantially earlier (Liu, Harris, & Kanwisher, 2002; Mouchetant-Rostaing, Giard, Delpuech, Echallier, & Pernier, 2000). While the interpretation of these differential effects is controversial (Johnson & Olshausen, 2003; VanRullen & Thorpe, 2001b), behavioural measurements avoid this problem.

Eye movements seem particularly well suited for measuring processing speed because some of them can be initiated in only 80–100 ms (Busettini, Masson, & Miles, 1997; Fischer & Weber, 1993; Masson, Rybarczyk, Castet, & Mestre, 2000). However, the visual processing required for such rapid responses is generally relatively simple. To demonstrate more sophisticated visual processing we need a task that requires participants to make some form of decision about the stimulus. Yet, even when natural images are used, oculomotor behaviour still appears to be governed by relatively low level characteristics of the scene (Parkhurst & Niebur, 2003). As a result, remarkably few studies have used eye movements to determine visual processing speed in higher level tasks (Gilchrist, Heywood, & Findlay, 2003; Levy-Schoen, Coeffe, & Jacobs, 1989; Pelz & Canosa, 2001).

☆ Statement: We declare that the work presented here has not been published elsewhere and is not under review with another journal. If published in *Vision Research* it will not be reprinted elsewhere in any language in the same form without the consent of the publisher, who holds the copyright.

* Corresponding author. Tel.: +33 5 62 17 28 02; fax: +33 5 62 17 28 09.
E-mail address: holle.kirchner@cerco.ups-tlse.fr (H. Kirchner).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.10.002

Nevertheless, it has been reported that in a go/no-go animal categorization task, participants can process two images presented in the left and right hemifield as quickly as when only one is present indicating parallel extraction of semantic object descriptions (Rousselet, Fabre-Thorpe, & Thorpe, 2002). We therefore hypothesized that if the differential neuromagnetic signal starting at 150 ms is a neural correlate of visual categorization (Rousselet et al., 2002; Thorpe et al., 1996), participants might be able to saccade to the side with the animal at around 180 ms after stimulus onset, assuming a delay for saccade preparation of 20–25 ms (Schiller & Kendall, 2004). However, in the present study, the fastest reliable eye movements were initiated after only 120 ms, implying that the visual system only needs roughly 95–100 ms to provide an initial first pass analysis of the images based on which a reliable behavioural response can be initiated. Furthermore, our results indicate that the differential ERP effects starting at 150 ms occur once initial stimulus processing has already been achieved.

2. Methods

2.1. Participants

Fifteen volunteers (mean age = 25 ± 3.5 years, 7 women and 8 men) with normal or corrected-to-normal vision performed a 2AFC visual discrimination task. The experimental procedures were authorized by the local ethical committee (CCPPRB No. 9614003). Experiments were undertaken with the understanding and written consent of each participant.

2.2. Experimental set-up

Participants were seated in a dimly lit room with their heads stabilized by a forehead and chin rest. Monochromatic natural scenes were presented on a video monitor (640 × 480, 100 Hz) on a black background at a distance of 80 cm resulting in an image size of 10° H × 14.5° V. The mean grey-levels of the target vs. distractor images were comparable.

Eight hundred and forty commercially available photographs were used, of which half were targets including a wide range of animals in their natural environments (mammals, birds, insects, reptiles, and fish); the remainder were distractors that included pictures of forests, mountains, flowers, and seascapes as well as man-made environments such as buildings and statues. Similar to a former study employing a go/no-go categorization task (Fabre-Thorpe et al., 2001), half of the images were seen only once, whereas the remainder were presented repeatedly. This manipulation allowed us to look for stimulus-specific learning effects.

2.3. Protocol

Two natural scenes were flashed for 20 ms centred at 6° in the left and right hemifield (see Fig. 1). The task was to make a saccade as fast as possible to the side where an animal had appeared. Targets were equiprobable in both hemifields. The fixation point disappeared after a pseudo-random time interval (800–1600 ms) leaving a 200 ms-time gap before the presentation of the images. This gap period generally serves to accelerate saccade initiation (Fischer & Weber, 1993; Saslow, 1967). After presentation of the images, two fixation crosses were presented for 1 s at ±6° to indicate the two possible saccade landing positions. The participants performed 10 blocks of 80 trials resulting in 200 trials per condition and participant (200 × 2 hemifields × new vs. repeated images = 800 trials).

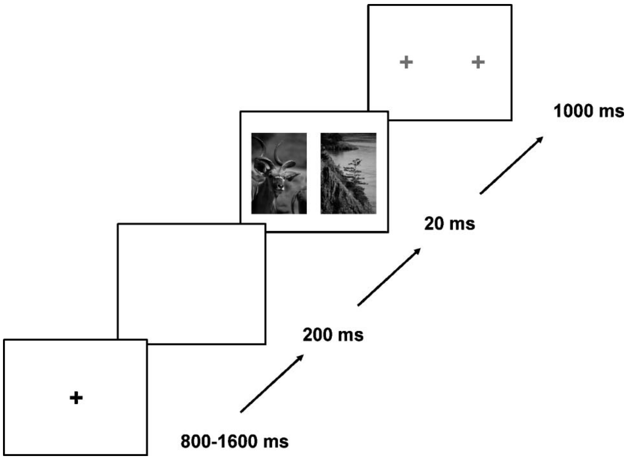

Fig. 1. Choice saccade task. After a pseudo-random fixation period, a blank screen (gap period) for 200 ms preceded the simultaneous presentation of two natural scenes in the left and right hemifields (20 ms). The images were followed by two grey fixation crosses indicating the saccade landing positions.

2.4. Response recording and detection

Eye position was recorded by horizontal EOG electrodes (1 kHz, lowpass at 90 Hz, notch at 50 Hz, baseline correction [−400:0] ms; NuAmps, Neuroscan) and stored on a PC. Saccadic reaction time (SRT) was determined off-line as the time difference between the onset of the images (time = 0) and the start of the saccade (see Fig. 2). As a first criterion, the difference signal between the left and right EOG electrodes had to exceed an amplitude threshold of ±30 μV (Fig. 2, black circle). Then, the saccade onset time was automatically determined as the nearest signal inflection preceding this point (Fig. 2, green circle). Each trial was verified by the experimenter to make sure that only the largest inflection (if any) was taken as a real saccade. Of the original 800 trials per participant, certain trials had to be rejected because of poor signal quality, with a mean rejection of 17% trials (*n* = 137) ranging from a minimum of 3% (*n* = 26) to a maximum of 34% (*n* = 271) (see Table 1).

To determine a value for the minimum SRT we divided the saccade latency distribution of each participant into 10 ms time bins (e.g., the 120 ms bin contains latencies from 115 to 124 ms) and searched for the first bin to contain significantly more correct than erroneous responses. This

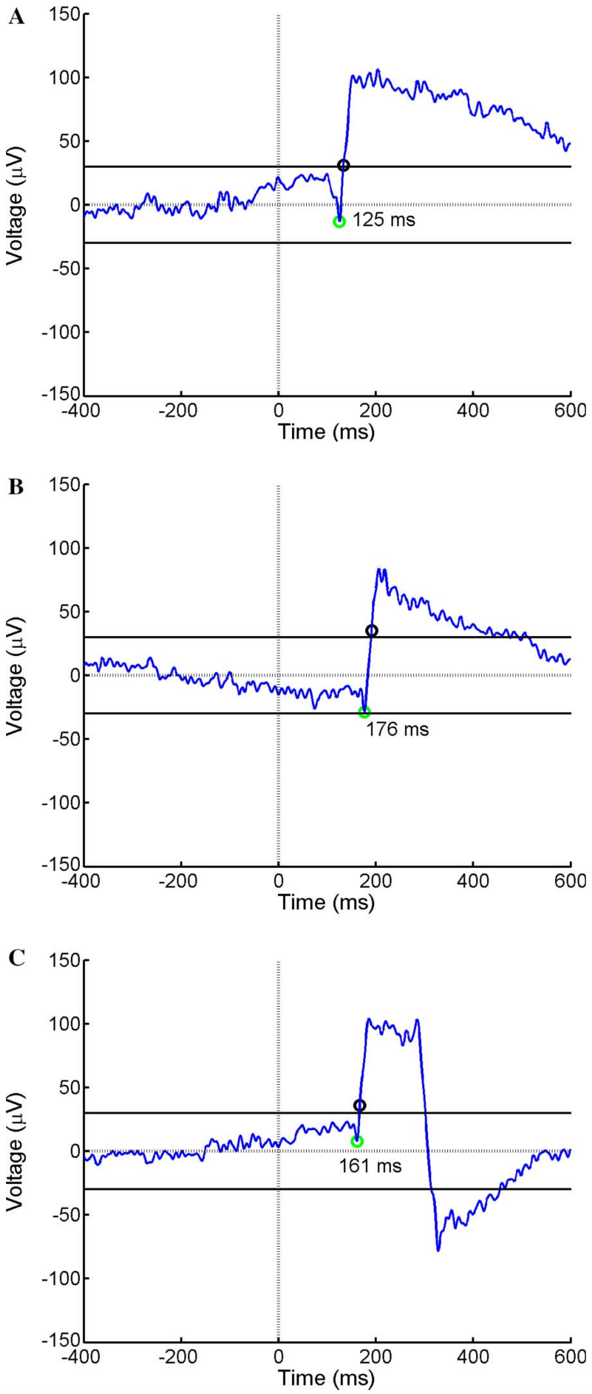

Fig. 2. Examples of saccade detections on a trial by trial basis using the horizontal EOG difference signals (in blue). Amplitude thresholds are indicated by dark horizontal lines. (A) Correct saccade to the left with a reaction time of 125 ms. (B) When a drift occurred prior to the first saccade, reaction time was determined as the nearest signal inflection (176 ms, green circle) preceding threshold crossing (black circle). (C) Erroneous response to the left which was corrected for soon after the first saccade; the response was counted as error with the reaction time of the first saccade (161 ms). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 1
Summary of behavioural results. Participant numbers correspond to those in Fig. 4

Subject	*N*	Accuracy (%)	Median RT (ms)	Min RT (ms)	Mean start (ms)
1	682	96.3	227	130	143
2	774	93.3	200	130	136
3	726	81.8	201	130	129
4	563	80.1	191	120	126
5	672	86.6	159	130	133
6	675	86.1	224	150	143
7	574	90.2	204	140	129
8	653	94.0	213	150	147
9	694	96.7	251	180	200
10	534	89.7	236	180	124
11	739	90.0	253	190	205
12	652	96.6	276	200	235
13	703	95.0	238	160	173
14	769	98.7	301	230	251
15	529	77.1	233	160	235
All	8998	90.1	228	120	140

The second column of this table indicates the total number of trials per participant (see Section 2 for details). Columns 3–5 give the mean accuracy, median and minimum reaction time values for each participant shown in Figs. 3B and C. The last column indicates the onset latency of the mean eye trace for each participant (see Fig. 5).

allowed us to eliminate trials involving anticipations which would result in chance performance (Kalesnykas & Hallett, 1987). χ^2 tests were calculated for each participant and bin. If at least 10 subsequent tests reached significance at the $p < 0.05$ level, the first of these bins was considered minimum SRT, or, if the participant made no errors in this latency range, the bin with minimum SRT had to contain at least five correct responses. This criterion was also applied in the control motor task.

3. Results

Despite the high demands made on the visual system by such a speeded visual discrimination task, performance was remarkably good (average = 90.1%) with one of the 15 participants achieving 98.7% correct responses. The median saccadic reaction time (SRT) in correct trials was 228 ms, although this value varied considerably between participants, from a minimum of 159 ms to a maximum of 301 ms (Fig. 3, Table 1).

The *minimum SRT*, i.e., the first 10 ms bin to contain significantly more correct than erroneous responses, across the whole data set was the 120 ms bin, corresponding to saccades initiated in the period 115–124 ms. Even when calculated individually, minimum SRTs were below or equal to 150 ms for eight of the participants (Fig. 3C in red, Table 1), and the mean minimum SRT across participants was 150 ms. The correlations between median reaction times and average accuracy ($r = 0.52$, Fig. 3B) and between minimum SRT and the accuracy level at minimum SRT ($r = 0.57$, Fig. 3C) indicated a speed-accuracy trade-off. However, even at minimum SRT average accuracy was 85.9% correct. This was no fluke, because in every participant a significant advantage of correct over erroneous responses continued for at least 120 ms after the minimum SRT. This can be seen in Fig. 4 which shows the latency distributions of correct and erroneous trials separately for each participant. The first column contains the 8 participants who obtained minimum SRTs below or equal to 150 ms. For many of these participants the latency distribution is clearly bimodal and the minimum SRT falls into the early peak of the distribution. The remaining 7 participants with minimum SRTs longer than 150 ms (second column of Fig. 4) all show a unimodal latency distribution. Irrespective of the form of the latency distribution, however, most of the erroneous responses occurred early in the latency range of each participant, and the number of errors rapidly decreased after minimum SRT. The minimum SRT thus indicates the point in time at which participants have access to sufficient visual information to perform the task successfully, because correct responses clearly outnumber errors after this point.

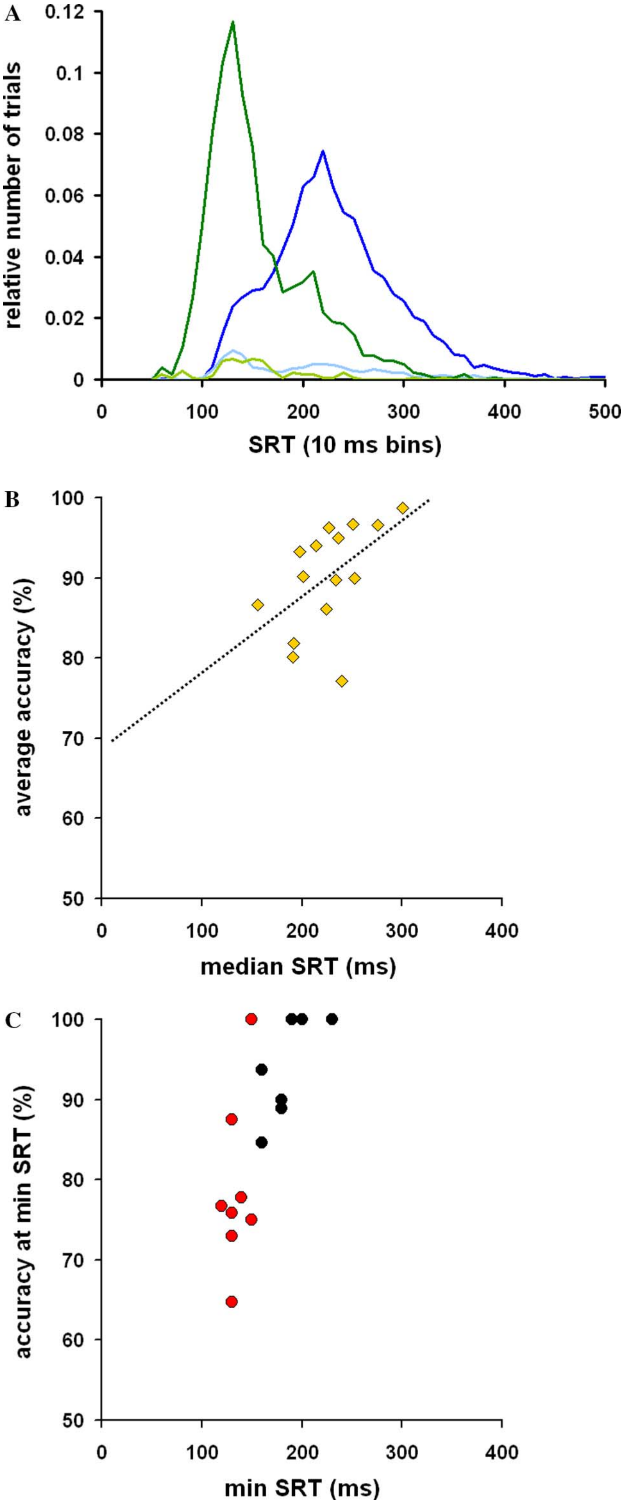

Fig. 3. Behavioural performance measures. (A) Saccadic reaction time (SRT) distributions in the choice saccade task (in blue) compared to the control motor task (in green). Correct responses are indicated in dark, direction errors in light. In the choice task, 15 participants saw two images in the left and right hemifield, and they were asked to make a saccade to the side containing an animal. In the control task, only a single image was presented either left or right of fixation, and eight of the participants were asked to make a saccade as soon as they detected the image. (B) Average accuracy as a function of median reaction time for each of the 15 participants in the choice task. The dashed line plots a linear regression and indicates the presence of a speed-accuracy trade-off ($r = 0.52$). (C) Accuracy at minimum reaction time for each of the 15 participants in the choice task. The eight fastest participants (in red) had minimum reaction times below or equal to 150 ms. Across participants, the minimum reaction time varied gradually between a minimum of 120 ms and a maximum of 230 ms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

All the former analyses are based on saccadic reaction times determined on a trial by trial basis. How do these compare with averaged eye traces, a method that is generally used in event-related potential studies and more recently in studies on fast ocular following responses (Masson et al., 2000) and the fast online control of pointing arm movements (Brenner & Smeets, 2004)? To calculate average eye traces in the present study, in a first step we determined the mean eye response for each participant separately. This was achieved by dividing trials into correct and erroneous responses when the target was either left or right of fixation. By averaging across correct and erroneous responses in the same direction (either leftward or rightward responses) we avoided any bias in favour of early correct responses. A final subtraction of mean leftward from mean rightward responses left us with the overall mean eye trace of each participant (see grey lines in Fig. 5). The onset of these individual mean responses are indicated in Table 1. As can be seen, for 9 of the 15 participants the onset of the mean response was less than 150 ms. In a final step, we determined the average eye trace across participants by averaging across the mean eye trace of each participant (black line in Fig. 5). Consecutive *t* tests on this grand average determined an overall onset latency of 140 ms (the interval between 135 and 144 ms). This latency is thus comparable to the mean minimum reaction time of 150 ms determined on a trial by trial basis (see Fig. 3C).

3.1. Motor-related control

The minimum SRTs and the onsets of the mean eye trace of each participant (see Table 1) are clearly very short. How would they compare with a simpler reaction time paradigm with no requirement to detect a particular

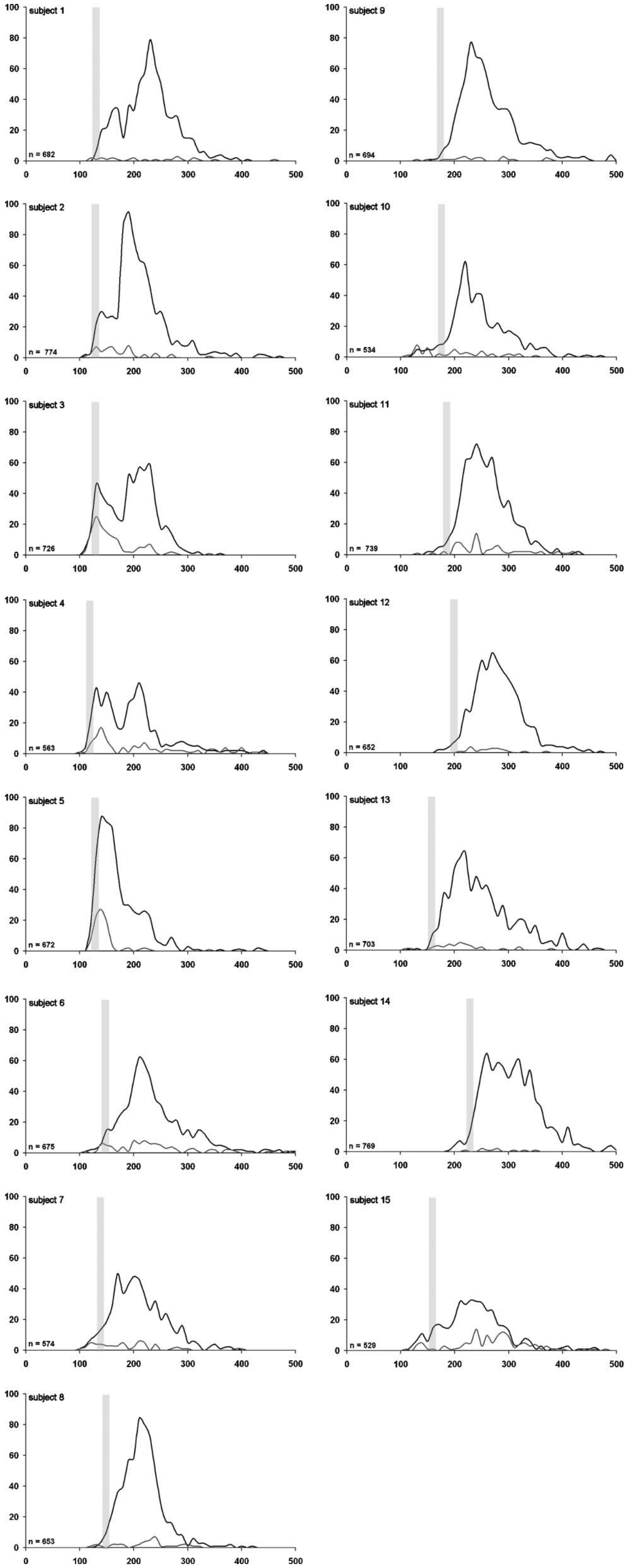

Fig. 4. Saccadic latency distributions of each participant. The abscissa shows the time after image onset (10 ms bins), and the ordinate shows the number of trials separated into correct responses (in black) and errors (in dark grey). The vertical, grey-shaded bar indicates the bin with the minimum reaction time for that participant. The numbers in the lower left corners give the total number of trials for each participant. The left column of the figure contains the eight participants who obtained minimum reaction times of 150 ms or less, and the right column of the figure shows the remaining seven participants with minimum reaction times longer than 150 ms.

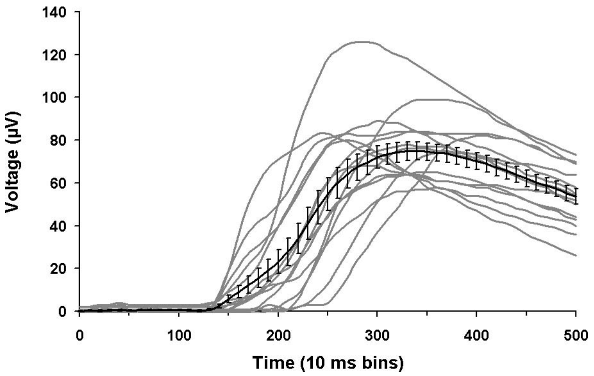

Fig. 5. Mean eye traces. Following the logic of recent studies on the fast online control of ocular following responses (Masson et al., 2000) and pointing arm movements (Brenner & Smeets, 2004), we determined the mean eye trace of each participant by subtracting leftward from rightward responses (see Section 3 for details). These individual traces are indicated in grey, and the grand average across these individual traces is indicated in black (±SEM). While for 9 of the 15 participants the onset latency of the mean trace was below 150 ms, the average eye trace across participants started to diverge from baseline in the 140 ms bin. This value is comparable to the mean minimum reaction time of 150 ms determined on a trial by trial basis (see Fig. 3C).

object category? To test this, eight of the participants performed a motor control task in which only a single image containing an animal was presented right or left of fixation. The participants were asked to make a saccade as soon as they detected the image. Their median SRTs (133 ms, Fig. 3A) were similar to those reported for human express saccades (120 ms, Fischer & Weber, 1993). Thus, a choice saccade can be initiated at a point where roughly half the responses in the motor control task have yet to be made. To give an estimate of the additional time needed to detect an animal in one out of two scenes presented in parallel we subtracted the minimum SRTs in the control motor task (90 ms, range 80–110 ms), from the minimum SRTs in the main choice saccade task. The average difference in minimum times between a simple oculomotor response and a more complex visual decision amounted to only 47.5 ms (range 20–80 ms).

3.2. Images triggering short latency saccades

What sorts of images can be processed at such extreme speed? If we took the complete set of images producing SRTs below 150 ms, many would correspond to errors, because most participants tended to make anticipations. This can be seen in the early peak of the latency distribution (Fig. 3A, in blue) and by the fact that about half of the participants had minimum SRTs longer than 150 ms (Fig. 3C, Table 1). However, two participants (participants 1 and 8 in Fig. 4) made virtually no errors in the first reliable latency bins (130–150 ms) allowing us to examine the set of images that resulted in these minimum SRTs (Fig. 6). Although some of the correctly detected animal pictures can be dismissed as being the result of

Fig. 6. Natural scenes triggering responses with reaction times between 130 and 150 ms in two of the fastest participants. In 51 trials, these participants together only made 5 errors (bottom line). Of the remaining 46 correct responses, two images were seen repeatedly, resulting in a total of 42 images. A further four images (3 distractors and 1 target) with reaction times between 119 and 128 ms are not shown, because they were presumably anticipations.

chance, the great majority of them must have been successfully processed even in this very short time period despite the fact that they vary considerably in size, viewpoint and type of animal.

3.3. Image-specific learning effects?

On half of the trials both the target and distractor images were seen only once by each participant, whereas for the remaining 400 trials we used a fixed set of 20 targets and 20 distractors that were presented in all their possible combinations. This allowed us to test whether familiarity improves performance. Both median SRTs and accuracy were virtually identical for novel and repeated images (226 vs. 228 ms; 89.5 vs. 90.6%), and no significant differences in accuracy were seen at any point of the reaction time distributions. Our results imply that image specific learning cannot improve processing speed since even novel images were responded to at the same optimal speed as images that had been seen several times, in line with an earlier report (Fabre-Thorpe et al., 2001). There was a tendency for SRTs to decrease during the course of the experiment (first vs. last block of 80 trials: 27 ms). However, this general learning effect did not interact with the type of image present (new vs. repeated) and was therefore probably due to improvement in motor skill.

3.4. Differences in processing speed between individual images?

The combination of 20 repeated targets and 20 repeated distractors resulted in a total of 300 trials per target image (20 trials × 15 participants). Thus we had enough statistical power to look for differences in processing speed between individual target images. For each of these targets, the reaction times were found to be normally distributed, but the means varied between a minimum of 215.5 ms and a maximum of 259.5 ms. The factor of target image was highly significant (ANOVA, $F_{(1,19)} = 5.96$, $p < 0.001$). In addition, the images with the shortest reaction times tended to be those with the highest accuracy (see Fig. 7). This negative correlation between mean reaction time and performance ($r = -0.71$) rules out a simple speed-accuracy trade-off explanation and implies that some animal targets can be processed faster and more efficiently than others.

Fig. 7. Performance measures for the 20 repeated target images. For each of 15 participants, the same twenty target images were each presented once with each of twenty repeated distractors. From left to right, the first number gives the mean reaction time (±SEM) in milliseconds, the second number gives the minimum SRT in milliseconds, and the third number gives the average accuracy in percent correct.

There were also significant variations in error rates between the 20 repeated distractor images, ranging from a minimum of 4% to a maximum of 20.6% (ANOVA, $F_{(1,19)} = 4.1$, $p < 0.002$). These error rates correlated positively with the mean reaction times of the correct responses to the side with the animal ($r = 0.82$), indicating that those distractors which induced the most errors also tended to slow responses to the target. Erroneous saccades to the side with the distractor had comparable mean reaction times across individual distractor images (see Fig. 8), and the correlation of error latencies with error rates was small ($r = -0.45$). Paired comparisons between reaction times for erroneous and correct responses separated by distractor image resulted in shorter error latencies (212 ± 6 ms) than correct responses (235 ± 2 ms), indicating that erroneous saccades to the side with a distractor image were likely to involve anticipations.

3.5. Correlation with basic image descriptors

Our participants were clearly able to perform well on the choice saccade task, even when they made very rapid eye movements with reaction times of 120–150 ms. Can we be sure that these responses were really dependent on the presence of an animal? One alternative explanation could be that the oculomotor system might have a natural bias towards some particular visual cue (for example, images with high local contrast), and that this particular image property just happened to be more prevalent in our animal images. Although our targets and distractors were matched for mean luminance, eliminating the possibility that participants could perform the task at above chance by simply making a saccade to the side with the brighter image, we did not match the images for all possible cues. Indeed, when we calculated a set of 13 first and second order statistics, as well as the fall-offs in the amplitude spectra, to analyze the

264±6 206±13 20.6
242±5 207±16 16.6
240±5 209±8 13.8
244±6 182±14 12.8
236±6 259±43 12.7

244±5 216±22 11.5
242±5 215±15 10.4
235±5 218±21 10.2
232±5 236±33 10.1
240±6 180±19 9.9

240±5 226±24 8.6
231±5 201±19 7.9
232±5 176±12 7.9
228±5 180±14 7.5
241±5 198±11 6.2

243±6 188±14 6.0
225±5 265±45 5.6
224±4 219±18 4.5
225±4 188±31 4.0
232±4 266±18 4.0

Fig. 8. The 20 repeated distractor images. From left to right, the first number gives the mean reaction time (±SEM) in milliseconds of responses in the direction of the targets, the second number gives the mean reaction time (±SEM) in milliseconds of responses in the direction of the distractor, and the third number gives the average error rate in percent correct.

target and distractor images, we found that many of these measures differed significantly between the two image categories. Table 2 provides the details of these analyses. Clearly then, we need to take seriously the possibility that a simple oculomotor bias could explain some of our results, a point that is particularly critical in the case of the fastest saccadic responses.

To test this possibility, we performed the following post hoc analyses: every time we found significant differences in the average value of some particular image descriptor between the target and distractor images, we removed 1 or 2 images with extreme values from the data pool to produce two new reduced sets of images that no longer differed with respect to the variable in question. In every case, we found that median reaction time and accuracy values calculated with the restricted sets were unchanged by eliminating the outliers. This is strong evidence that none of these differences could reliably be used by the participants to perform the task. However, even if we have ruled out the possibility that a simple bias in the oculomotor system for one of these low level image properties could explain participants' performance, one could still argue that some *combination* of these low-level image statistics might be involved. For example, one might propose that participants tend to make saccades to images that had both high contrast and high second-order homogeneity (or some other combination). However, even this possibility can be excluded. In one final manipulation, we removed 10 targets and 10 distractors from the test set to produce two populations of images that showed no differences in any of the 13 different first and second order statistics nor in the spectral amplitude fall-offs we used. Performance with the remaining 10 targets and distractors was again identical to that seen with the complete data set (median = 230 ms, accuracy = 90.5% correct).

Table 2
Summary of statistical tests between target and distractor images with regard to low-level image descriptors

		Mean	*SE*	*t* test	*df*	$p < 0.05$	*N*	Median	Accuracy
First order									
Mean	Target	91.20	1.42	−0.29	38.0	n.s.			
	Distractor	90.60	1.49						
Variance	Target	2239.42	221.32	2.83	32.7	0.008	2	230	90.5
	Distractor	3382.62	338.39						
SNR	Target	2.06	0.10	−2.04	38.0	0.048	1	229	90.8
	Distractor	1.73	0.12						
RMS	Target	0.51	0.03	2.64	38.0	0.012	2	229	90.5
	Distractor	0.63	0.04						
Skewness	Target	0.36	0.11	0.61	38.0	n.s.			
	Distractor	0.46	0.12						
Kurtosis	Target	3.07	0.19	−0.64	38.0	n.s.			
	Distractor	2.90	0.20						
Second order									
Energy	Target	0.0006	0.0006	2.68	28.9	0.012	2	229	91.0
	Distractor	0.0013	0.0011						
Contrast	Target	321.57	176.89	3.12	29.1	0.004	3	230	90.9
	Distractor	582.70	329.66						
Entropy	Target	8.48	0.54	0.16	38.0	n.s.			
	Distractor	8.51	0.66						
Homogen	Target	0.17	0.06	1.41	38.0	n.s.			
	Distractor	0.21	0.08						
Correlation	Target	0.49	0.13	2.01	38.0	0.052	1	229	90.8
	Distractor	0.59	0.16						
Spectral amplitude fall-off									
Horizontal	Target	1.43	0.03	2.85	38.0	0.007	2	229	90.8
	Distractor	1.27	0.05						
Vertical	Target	1.46	0.04	2.9	38.0	0.006	2	229	91.1
	Distractor	1.31	0.04						
Combined reduction							10	230	90.2

First order image statistics are based on the distribution of grey-scale values in the images: mean, variance, signal to noise ratio (SNR), root mean square (RMS), skewness and kurtosis. Second order image statistics such as energy, contrast, entropy, homogeneity (homogen), and correlation are based on the means across co-occurrence matrices of grey-scale values for adjacent pixels in four different directions within the images ($\theta = 0°$, 45°, 90°, and 135°; Baraldi and Parmiggiani, 1995). Indices of spectral amplitude fall-offs were calculated by curve-fits to the horizontal and vertical components of the centred Fourier-transform of the images (van der Schaaf & van Hateren, 1996). The first column gives the mean value in each image category (target vs. distractor), followed by the standard error of the mean (SE). The critical t value, its degrees of freedom (*df*) and its exact probability ($p < 0.05$) indicate whether targets and distractors differed significantly on this image dimension. In such cases, we removed a given number (Ncut) of images with extreme values from each category and determined the median reaction time (median in ms) and average accuracy (accuracy in % correct) in the reduced data set. The last line (combined reduction) shows that behavioural performance in the reduced data set (when all the images with extreme values for any of the image descriptors have been removed) was still almost identical to the overall performance level (median = 228 ms, accuracy = 90.1%)

Together, these findings provide strong evidence that our participants' ability to perform the task can not be explained by spurious differences between the target and distractor images that just happened to correlate with a natural tendency of the oculomotor system to initiate saccades towards particular types of images. It also argues against the hypothesis that participants could develop a strategy for detecting animals in natural images simply by using these low level descriptors or their combination.

Nevertheless, while we can exclude the possibility that these low level descriptors could be a sufficient cue to perform the task, this does not mean that they are irrelevant. Indeed, when we took the 20 repeated targets and looked for image factors that might be correlated with high performance (low minimum RT, high accuracy, or both), we found a number of interesting correlations. In particular, short mean reaction times were associated with targets containing high SNR, high kurtosis, high energy, low entropy, and small correlation (see Table 3). Likewise, high accuracy was associated with targets containing low levels of variation in the grey-scale values and high SNR. This implies that the most rapidly and accurately processed targets were the most homogeneous and texturally uniform, a suggestion that is born out by an inspection of the images in Fig. 7. Note, for example, that the most efficiently detected animal target (top left corner) is a remarkably homogenous image (high SNR, high energy, and low variance), whereas the less efficiently detected targets include several images with a large variability in the grey-scale values (low SNR, low contrast).

The situation in the case of the distractor images was quite different. Here we found that none of the first and second order image statistics nor the spectral amplitude fall-offs showed any tendency to correlate with either error rates or short latency responses in the direction of the targets. This is again evidence that participants could not decide to make a saccade in a particular direction on the basis of these image statistics alone. On the other hand, it could be that in the presence of diagnostic animal features, certain combinations of image statistics might allow saccadic responses to be made more efficiently. This would explain why particular target images had both abnormally short reaction times and high accuracy.

4. Discussion

Until now there has been little evidence in the literature that rapid eye movements are influenced by the semantic content of a scene (Findlay & Gilchrist, 2003; Henderson & Hollingworth, 1999). However, using a choice saccade task we show here that human participants can reliably detect an animal in a natural scene and initiate the appropriate saccade in as little as 120 ms. These unexpected results have a number of implications that will be discussed in the following paragraphs.

4.1. Relation to express saccades

Inspection of individual latency distributions indicated that several of those participants with particularly short minimum reaction times showed bimodality in their latency distributions. In previous studies using a similar gap paradigm to the one used here, the presence of such a separate, early peak in the latency distribution has been taken as evidence to define a separate class of "express" saccades (Fischer & Weber, 1993, but see also Kingstone & Klein, 1993; Wenban-Smith & Findlay, 1991). However, it is

Table 3
Correlation of behavioural performance of the 20 repeated targets with basic image descriptors

	Mean	Min	Max	Corr RT	$p < 0.05$	Corr acc	$p < 0.05$
First order							
Mean	91.2	78.92	99.13	0.07	n.s.	−0.29	n.s.
Variance	2239.42	938.3	4945.78	0.41	n.s.	−0.54	0.01
SNR	2.06	1.33	3.1	−0.46	0.04	0.45	0.04
RMS	0.51	0.32	0.75	0.42	n.s.	−0.44	0.05
Skewness	0.36	−0.35	1.19	0.09	n.s.	−0.10	n.s.
Kurtosis	3.07	1.83	5.39	−0.45	0.05	0.36	n.s.
Second order							
Energy	0.0006	0.0002	0.0027	−0.57	0.01	0.36	n.s.
Contrast	321.57	101.52	899.15	0.15	n.s.	−0.33	n.s.
Entropy	8.48	7.26	9.25	0.58	0.01	−0.37	n.s.
Homogen	0.17	0.09	0.30	−0.29	n.s.	0.07	n.s.
Correlation	0.49	0.24	0.73	0.44	0.05	−0.43	n.s.
Spectral amplitude fall-off							
Horizontal	1.4	1.2	1.8	−0.22	n.s.	−0.28	n.s.
Vertical	1.5	1.2	1.7	0.10	n.s.	−0.27	n.s.

First and second order image statistics are based on the distribution of grey-scale values in the images, and indices of spectral amplitude fall-offs are based on the horizontal and vertical spatial frequency components of the images (see caption of Table 2 for details). The first three columns indicate the mean, minimum (min), and maximum (max) values across all 20 repeated target images for each image descriptor. The last four columns show the coefficients of Pearson's bivariate correlations (corr) between these image descriptors and mean reaction times (RT) or log-scaled average accuracy (acc) together with the exact probability ($p < 0.05$). As can be seen, short mean reaction times were associated with targets containing high SNR, high kurtosis, high energy, low entropy, and small correlation, and high accuracy was associated with targets containing relatively little variation in the grey-scale values and high SNR.

interesting to note that some of the participants (participants 6–8 in Fig. 4) had very short minimum reaction times in the range of 140–150 ms without any evidence for bimodality in their latency distribution. Thus we believe that short minimum reaction times are not necessarily associated with express saccades, at least when the latter are defined on the basis of bimodality.

4.2. Underlying neural circuits

If we take the minimum reaction time across all conditions and participants which amounted to 120 ms, and we assume a delay of about 20–25 ms for the saccade-generator in the brainstem to program the response (Schiller & Kendall, 2004), the underlying visual processing needs only about 95–100 ms to be effective. One obvious problem is to understand which anatomical pathways might be used to initiate such rapid oculomotor responses to complex natural scenes. One possibility could involve only subcortical processing via the superior colliculus, but the complexity of the natural scenes that we used here makes this hypothesis rather unlikely. There are suggestions of a subcortical route to the amygdala involved in responses to fear-inducing visual stimuli, such as spiders or snakes (Adolphs et al., 2005; Morris, Ohman, & Dolan, 1999). But, again, the wide range of animal targets that produced short latency responses and that had no obvious threat value for humans would be a problem for such an explanation.

Current evidence instead points to a cortical route involving the ventral visual pathway implicated in object recognition (Chelazzi, 1995; Girard, Lomber, & Bullier, 2002; Grill-Spector, 2003; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Logothetis, 1998; Perrett, Hietanen, Oram, & Benson, 1992; Rainer, Lee, & Logothetis, 2004; Rolls, 2000; Sheinberg & Logothetis, 2001; Sugase, Yamane, Ueno, & Kawano, 1999; Tanaka, 1993). With respect to saccadic eye movements, it is likely that the task involves the frontal eye fields (FEFs) and the lateral intraparietal sulcus (LIP) of the posterior parietal cortex that are both known to receive substantial projections from a range of areas in the ventral visual pathway (see Fig. 9). On the cortical level, the FEFs and LIPs are the main structures involved in the visual selection of the target location and response preparation (Pare & Wurtz, 2001; Shadlen & Newsome, 2001; Thompson, Hanes, Bichot, & Schall, 1996). The decision when to initiate a saccade appears to be taken when the level of activity in the motor neurons reaches a fixed threshold, and the variability in response time can be accounted for by the time taken by these motor neurons to reach the threshold (Hanes & Schall, 1996). So it could be that in our task cortical eye fields, such as FEFs and LIPs, in the left and right hemispheres independently accumulate evidence from ventral stream areas that favour the presence of an animal in the contralateral visual field.

This hypothesis is corroborated by observations from studies using transcranial magnetic stimulation (TMS) over the human FEF which indicate that target discrimination processes in difficult search tasks can occur very early, i.e., as early as 40–80 ms after display onset (Juan & Walsh, 2003; Nyffeler et al., 2004; O'Shea, Muggleton, Cowey, & Walsh, 2004), even if the participants perform the task covertly, that is, without preparing a saccadic response (Juan, Shorter-Jacobi, & Schall, 2004). Converging evidence from electrophysiology (Thompson & Bichot, 2005) and TMS (Juan et al., 2004; O'Shea et al., 2004) thus indicates that cortical eye fields, such as the FEFs, are not only implicated in saccade programming, but they also contain visually responsive neurons which might accumulate evidence from other visual cortical areas at extremely high speed.

The latency values in Fig. 9 are based on single-unit recording data from macaque monkeys. The corresponding latencies in humans are probably longer (Fabre-Thorpe, Richard, & Thorpe, 1998; Mace, Richard, Delorme, & Fabre-Thorpe, 2005, but see also Michel, Seeck, & Murray, 2004). If we assume that the cortical eye fields accumulate evidence from ventral visual areas, this latency data indicates that performance in our forced-choice saccade task might be based on information coming from mid- and low-level ventral areas, possibly no higher than area V4. Rapid, parallel information update between areas V4 and the cortical eye fields, and/or a direct route from V4 into the deep layers of the superior colliculus (Fries, 1984; Webster, Bachevalier, & Ungerleider, 1993), might be one of only a few routes sufficiently rapid to trigger these extremely fast saccadic responses (dark arrows in Fig. 7). Could it be that fast animal detection in natural scenes bypasses higher level visual and/or motor areas, such as IT and/or the cortical eye fields? This clearly is a rather provocative hypothesis that merits further investigation.

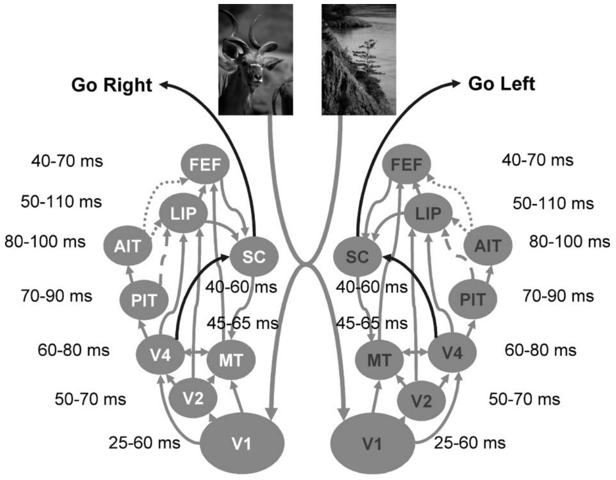

Fig. 9. Flow-chart of anatomical connections between the ventral visual pathway involved in object recognition and the eye fields involved in eye movement control. Latency estimates are based on monkey data, but the corresponding values in humans are probably longer. The first number for each processing stage indicates the approximate latency of the earliest responses of visually responsive neurons to a flashed stimulus, whereas the second number provides a more typical, average latency (Nowak & Bullier, 1997; Thorpe & Fabre-Thorpe, 2001). Arrows from the posterior and anterior inferior temporal cortex are dashed and dotted, because these connections are either insignificant or absent (Bullier et al., 1996; Schall et al., 1995).

Irrespective of the precise neural pathways involved, the current results leave little room for time-consuming iterative processing. We should stress that it is not because the eyes can move in the correct direction in just 120 ms after scene onset that vision has been completed at this time. It is likely that the first wave of processing, while sufficient to trigger useful behavioural responses, is just the start of a series of complex events involving feedback loops to earlier visual areas (Foxe & Simpson, 2002; Hamker, 2003; Moore & Armstrong, 2003). These secondary processes might be required for segmenting the scene, selecting the part of the image where the target is located and generating conscious perception.

4.3. Minimum reaction time

Note that our interpretations of current neurobiological data are based on the outcome of our minimum reaction time measure. While the latter has already been used in a number of previous studies (Rousselet et al., 2002; Thorpe et al., 1996; VanRullen & Thorpe, 2001a) it still is controversial. There are, however, a number of reasons for establishing such a measure when the aim is to determine sensory processing speed. As already mentioned, converging evidence from electrophysiology and TMS indicates that structures responsible for saccade programming, such as the FEFs, contain not only motor neurons, but also visually responsive neurons, the functions of which can be clearly distinguished. FEF *movement neurons* do not respond to visual stimulation, but fire before and during saccades, signalling whether and when to make a saccade (Hanes & Schall, 1996). In contrast, *visually responsive neurons* in FEF perform target discriminations independently of saccade programming (Juan et al., 2004), and the timing of the discrimination process does not predict the variability of saccadic reaction times (Murthy, Thompson, & Schall, 2001). This distinction into motor-related and visually responsive neurons has been further clarified recently by electrophysiological recordings in monkey IT and FEF in a saccadic choice reaction time task (DiCarlo & Maunsell, 2005). By taking into account mean neuronal latencies and the trial by trial covariance between these neuronal latencies and the mean behavioural reaction times of the monkeys, it can be shown that visually responsive neurons in IT and FEF have little covariance with behavioural reaction times, whereas motor-related neurons in FEF covary tightly with the behavioural response (DiCarlo & Maunsell, 2005). With regard to the interpretation of behavioural data in humans, such as in the present study, this means that most of the variability in the latency distribution should be considered motor. This in turn implies that the variability in reaction times cannot be used to deduce directly the underlying visual response time (Ratcliff & Smith, 2004). There is a parallel with recent work on express saccades which indicates that the presence or absence of a bimodal reaction time distribution is less a result of differences in the sensory pathways involved, but seems more a question of the state of the motor system in a particular sequence of trials (Carpenter, 2001).

Finally, we note that while eight of our 15 participants had minimum reaction times of 150 ms or less, for the remaining seven participants minimum reaction times could be substantially longer (minimum reaction times varied gradually from a minimum of 120 ms to a maximum of 230 ms). It seems likely that our inability to obtain shorter values for the slower, but more accurate participants (speed-accuracy trade-off) is related to differences at the decision and motor stages, but probably not to differences in the underlying speed of sensory processing. When participants take longer to respond, they can accumulate more information about the stimulus, probably involving iterative processing, thus allowing for higher levels of accuracy. For this reason, it is useful to analyse minimum reaction time for each participant individually. However, if we are interested in the speed of initial sensory analysis corresponding to the first wave of processing (VanRullen & Thorpe, 2002), we would argue that only the minimum reaction times of the fastest participants are really critical.

Of course, the actual value of the minimum reaction time measure directly depends on the total number of trials per participant—with more trials, the minimum reaction time might be even shorter. However, given the very short values reported here, it seems likely that we are close to the limits imposed by the conduction velocities of the neurons involved. As a result, the minimum reaction times reported here should be considered a valid estimate of the time available.

4.4. Relation to EEG and MEG studies of processing speed

The present results are also very informative in the light of the recent debate about the interpretation of short latency differential EEG and MEG signals. Earlier reports from our group suggested that the differential response starting at 150 ms after scene onset reflects high-level categorization processes (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Rousselet et al., 2002; Thorpe et al., 1996; VanRullen & Thorpe, 2001b). In one particular study, VanRullen and Thorpe discussed the fact that the differential effects occurring earlier than 150 ms might be related to low-level physical differences irrespective of the meaning of the stimuli, but by inversing the target category in a go/no-go task these authors showed that the differences arising at 150 ms or later could not be explained solely on the basis of such low-level cues (VanRullen & Thorpe, 2001b). However, Johnson and Olshausen recently relaunched this issue by suggesting that even the 150 ms differential effect might relate only to low-level differences between images, and that only those differential effects that co-vary with reaction time should be considered as relevant for categorization (Johnson & Olshausen, 2003, 2005). However, there can be no doubt about the significance of the short latency behavioural responses reported here. Indeed, Fig. 5 shows effectively the results of the same sort of averaging procedure that has previously been used for analyzing ERP responses (Rousselet et al., 2002; Thorpe et al., 1996), but here analyzing the eye trace signals. As can be seen, the onset of the average eye trace starts to diverge from baseline at 140 ms, and for nine of the 15 participants these onsets were shorter than 150 ms (Table 1). Thus, since the eyes can move even before the onset of the 150 ms visual differential response, it will be very important to re-examine some of the earlier differential effects that have been described previously in a variety of other tasks (Liu et al., 2002; Mouchetant-Rostaing et al., 2000), but neglected because they appeared too early to be pertinent to high level categorization processes.

Based on the present results, our hypothesis is that the differential visual effects starting at 150 ms may not, in fact, reflect the initial wave of processing required to find an animal in a scene, but rather may be related to secondary, re-entrant selection processes that occur after the first wave of stimulus processing has been completed.

4.5. Implications for computational models

A further key issue is the nature of the visual computations that allow these very fast decisions to be made. There have been suggestions that some form of scene categorization can be achieved by analyzing the energy distribution across a set of orientation and spatial frequency tuned channels, something that could be done early on in the visual system. For example, Torralba and Oliva found that a model based on a set of such orientation and spatial frequency tuned channels achieved accuracy rates of 80–85% on image categorization problems such as “city,” “interior,” “beach,” or “mountain” (Torralba & Oliva, 2003). Could such a strategy be used by our participants to initiate rapid saccadic eye movements toward the image containing an animal? And, relatedly, could it be that our results come from spurious differences between the target and distractor images which just happen to correlate with a tendency of the oculomotor system to initiate saccades towards particular types of images, independently of whether there was an animal present or not? To address these questions, we took advantage of a particular feature of the current experimental design. Our choice of presenting a fixed number of target and distractor images in all their possible combinations to a large number of participants produces a rich dataset that can be very useful for testing specific hypotheses about how the participants might be performing the task. First, we can look for image features that differ between the target and distractor images and see whether the presence of these features or combinations of them correlate with behavioural performance. Then, by excluding targets and distractors with extreme values we were able to generate reduced image subsets that were matched on any of these candidate measures. So far, our results have failed to find any evidence for the use of such low level descriptors because none of the reduced image sets produced changes in either average response time or accuracy. This includes strategies based on the amount of energy in the horizontal and vertical orientation channels (Torralba & Oliva, 2003).

Nevertheless, we would not claim that the set of image statistics used here has exhausted the list of low-level image characteristics that could potentially be involved in performing the task. The critical test would be to see whether, when a particular classification strategy has been developed, selecting a subset of targets and distractors that cancelled differences between the two image categories leads to a collapse in performance. Given that the present study provides full details of human performance with a wide range of images, it will now be possible to directly compare behavioural data with the outcome of various computational models of object and scene categorization. This is clearly an important topic for future work.

Although we have so far found little evidence that a simple strategy based on low-level cues could be sufficient to perform the task, the speed with which an animal is detected in the scene implies that the underlying processing must be relatively simple. A number of recently proposed biologically inspired computer vision models have the advantage that they can be implemented in feed-forward architectures (Riesenhuber & Poggio, 2002). One class of such image processing algorithms looks for particular combinations of diagnostic elements such as eyes or ears (Thorpe, Delorme, & Van Rullen, 2001). Other approaches have used information maximization to determine such diagnostic image fragments (Ullman, Vidal-Naquet, & Sali, 2002).

4.6. Relation to other experimental paradigms

Intriguingly, in a previous study, using a manual go/no-go categorization task, no evidence was seen that particular images were associated with abnormally short reaction times, although some images were clearly more difficult than others (Fabre-Thorpe et al., 2001). Furthermore, the images used in the present study all had been categorized perfectly in that previous manual go/no-go task. In contrast, in the current study, we found clear evidence that some targets could be processed faster and more efficiently than others. Why should there be a difference in results? One possibility is that go/no-go and choice tasks generally involve different sensory processing or the fact that, in the present experiment, each image was tested a large number of times. However, yet another possibility is that in our speeded choice saccade task, eye movements can be initiated before visual processing has been completed. This would explain why, although some images were apparently correctly categorized even when the eyes started to move at 120 ms, for other images, correct performance was only possible when the onset of the saccade was delayed. This would account for the striking correlation between reaction times and accuracy levels for the 20 repeated targets. In contrast, in the previous manual go/no-go task, the earliest motor responses did not start until at least 230–250 ms after stimulus onset, and it might therefore be that by this time sensory analysis of all the images had been completed, so that all images appeared to be equally rapidly processed. Additional work will be needed to determine the reasons for the differences between the present results and the previous go/no-go study. However, the fact that the present choice saccade task allows us to directly determine processing speed for individual images, means that it should be possible to obtain detailed information about the kinds of objects and features leading to processing advantages in natural scenes. Such knowledge should allow us to test particular theories about object detection and categorization.

Finally, it is worth speculating about the reasons why this sort of very fast visual discrimination has not been reported previously. The key may lie in our use of a simple left-right decision. This could allow exceptionally fast responses because the oculomotor system can program both movements in advance (McPeek, Skavenski, & Nakayama, 2000). This simple choice constitutes a rather special case of active visual search tasks which generally include a whole array of possible target locations. In these more complex conditions, macaque monkeys exhibit minimum saccadic reaction times of about 200 ms (Schall, 2002), much longer than in our left/right forced choice task.

5. Conclusions

The experimental protocol used here seems very promising as a tool for determining processing speed. Similar tasks could easily be designed to measure the time required to extract information at different levels of complexity, from simple visual properties of isolated stimuli to complex scenes requiring categorical and semantic knowledge. The resulting data might help to build up a detailed picture of the temporal sequences involved in sensory analysis by determining those features and stimuli that are processed early in time. Already on the basis of the current results we can conclude that object related visual information can be extracted extremely rapidly from complex natural scenes and that, furthermore, this information can be directly used to control behaviour.

Acknowledgments

This work was supported by the European Research Training Network ‘Perception for Recognition and Action’ (HPRN-CT-2002-00226) and the Integrative and Computational Neuroscience Initiative of CNRS. We would like to thank Rufin VanRullen for insightful discussions during the course of the study. Pascal Girard made the calculation of first and second order image statistics accessible, and two anonymous reviewers made very helpful comments on an earlier version of the manuscript.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68–72.

Baraldi, A., & Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing, 33*(2), 293–304.

Brenner, E., & Smeets, J. B. (2004). Colour vision can contribute to fast corrections of arm movements. *Experimental Brain Research, 158*(3), 302–307.

Bullier, J., Schall, J. D., & Morel, A. (1996). Functional streams in occipito-frontal connections in the monkey. *Behavioural Brain Research, 76*(1–2), 89–97.

Busettini, C., Masson, G. S., & Miles, F. A. (1997). Radial optic flow induces vergence eye movements with ultra-short latencies. *Nature, 390*(6659), 512–515.

Carpenter, R. H. (2001). Express saccades: Is bimodality a result of the order of stimulus presentation? *Vision Research, 41*(9), 1145–1151.

Chelazzi, L. (1995). Neural mechanisms for stimulus selection in cortical areas of the macaque subserving object vision. *Behavioural Brain Research, 71*, 1–2.

Delorme, A., Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2004). Interaction of top–down and bottom–up processing in the fast visual analysis of natural scenes. *Brain Research Cognitive Brain Research, 19*(2), 103–113.

DiCarlo, J. J., & Maunsell, J. H. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *Journal of Neurophysiology, 93*(5), 2974–2986.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. J. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13*(2), 1–10.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport, 9*(2), 303–308.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision. The Psychology of looking and seeing*. Oxford: Oxford University Press.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16*, 553–610.

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.

Fries, W. (1984). Cortical projections to the superior colliculus in the macaque monkey: A retrograde study using horseradish peroxidase. *The Journal of Comparative Neurology, 230*(1), 55–76.

Gilchrist, I. D., Heywood, C. A., & Findlay, J. M. (2003). Visual sensitivity in search tasks depends on the response requirement. *Spatial Vision, 16*(3-4), 277–293.

Girard, P., Lomber, S. G., & Bullier, J. (2002). Shape discrimination deficits during reversible deactivation of area V4 in the macaque monkey. *Cerebral Cortex, 12*(11), 1146–1156.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology, 13*(2), 159–166.

Hamker, F. H. (2003). The reentry hypothesis: Linking eye movements to visual perception. *Journal of Vision, 3*(11), 808–816.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*(5286), 427–430.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 96*(16), 9379–9384.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3*(7), 1–13.

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision, 5*(4), 299–312.

Juan, C. H., Shorter-Jacobi, S. M., & Schall, J. D. (2004). Dissociation of spatial attention and saccade preparation. *Proceedings of the National Academy of Sciences of the United States of America, 101*(43), 15541–15544.

Juan, C. H., & Walsh, V. (2003). Feedback to V1: A reverse hierarchy in vision. *Experimental Brain Research, 150*(2), 259–263.

Kalesnykas, R. P., & Hallett, P. E. (1987). The differentiation of visually guided and anticipatory saccades in gap and overlap paradigms. *Experimental Brain Research, 68*(1), 115–121.

Kingstone, A., & Klein, R. M. (1993). What are human express saccades? *Perception & Psychophysics, 54*(2), 260–273.

Levy-Schoen, A., Coeffe, C., & Jacobs, A. M. (1989). Sensory factors are insufficient to define the ocular saccade goal in complex visual fields. *Brain, Behavior, and Evolution, 33*(2-3), 80–84.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*(9), 910–916.

Logothetis, N. (1998). Object vision and visual awareness. *Current Opinion in Neurobiology, 8*(4), 536–544.

Mace, M. J., Richard, G., Delorme, A., & Fabre-Thorpe, M. (2005). Rapid categorization of natural scenes in monkeys: Target predictability and processing speed. *Neuroreport, 16*(4), 349–354.

Masson, G. S., Rybarczyk, Y., Castet, E., & Mestre, D. R. (2000). Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. *Visual Neuroscience, 17*(5), 753–767.

McPeek, R. M., Skavenski, A. A., & Nakayama, K. (2000). Concurrent processing of saccades in visual search. *Vision Research, 40*(18), 2499–2516.

Michel, C. M., Seeck, M., & Murray, M. M. (2004). The speed of visual cognition. *Advances in clinical neurophysiology. Supplements to Clinical Neurophysiology, 57*, 613–623.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*(6921), 370–373.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences of the United States of America, 96*(4), 1680–1685.

Mouchetant-Rostaing, Y., Giard, M. H., Delpuech, C., Echallier, J. F., & Pernier, J. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *Neuroreport, 11*(11), 2521–2525.

Murthy, A., Thompson, K. G., & Schall, J. D. (2001). Dynamic dissociation of visual selection from saccade programming in frontal eye field. *Journal of Neurophysiology, 86*(5), 2634–2637.

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* (12, pp. 205–241). New York: Plenum Press.

Nyffeler, T., Bucher, O., Pflugshaupt, T., Von Wartburg, R., Wurtz, P., Hess, C. W., et al. (2004). Single-pulse transcranial magnetic stimulation over the frontal eye field can facilitate and inhibit saccade triggering. *The European Journal of Neuroscience, 20*(8), 2240–2244.

O'Shea, J., Muggleton, N. G., Cowey, A., & Walsh, V. (2004). Timing of target discrimination in human frontal eye fields. *Journal of Cognitive Neuroscience, 16*(6), 1060–1067.

Pare, M., & Wurtz, R. H. (2001). Progression in neuronal processing for saccadic eye movements from parietal cortex area lip to superior colliculus. *Journal of Neurophysiology, 85*(6), 2545–2562.

Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*(2), 125–154.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research, 41*(25-26), 3587–3596.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society London B Biological Sciences, 335*(1273), 23–30.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory, 2*(5), 509–522.

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biology, 2*(2), 275–283.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333–367.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12*(2), 162–168.

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron, 27*(2), 205–218.

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.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57*(8), 1030–1033.

Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society London B Biological Sciences, 357*(1424), 1073–1082.

Schall, J. D., Morel, A., King, D. J., & Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: Convergence and segregation of processing streams. *Journal of Neuroscience, 15*(6), 4464–4487.

Schiller, P. H., & Kendall, J. (2004). Temporal factors in target selection with saccadic eye movements. *Experimental Brain Research, 154*(2), 154–159.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86*(4), 1916–1936.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience, 21*(4), 1340–1350.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400*(6747), 869–873.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262*(5134), 685–688.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Progress in Brain Research, 147*, 251–262.

Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology, 76*(6), 4040–4055.

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

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Neuroscience seeking categories in the brain. *Science, 291*, 5502.

Thorpe, S. J., & Fabre-Thorpe, M. (2002). Fast visual processing and its implications. In M. Arbib (Vol. Ed.), *The handbook of brain theory and neural networks* second Edition, pp. 441–444). Cambridge, MA: MIT Press.

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

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

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7), 682–687.

van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research, 36*(17), 2759–2770.

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). The time course of visual processing: From early perception to decision- making. *Journal of Cognitive Neuroscience, 13*(4), 454–461.

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

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1993). Subcortical connections of inferior temporal areas TE and TEO in macaque monkeys. *The Journal of Comparative Neurology, 335*(1), 73–91.

Wenban-Smith, M. G., & Findlay, J. M. (1991). Express saccades: Is there a separate population in humans? *Experimental Brain Research, 87*, 1.