Animals roll around the clock: The rotation invariance of ultrarapid visual processing

Rudy Guyonneau Centre de Recherche "Cerveau et Cognition," UMR 5549, Toulouse, France

Holle Kirchner Centre de Recherche "Cerveau et Cognition," UMR 5549, Toulouse, France

Simon J. Thorpe Centre de Recherche "Cerveau et Cognition," UMR 5549, Toulouse, France

The processing required to categorize faces and animals is not only rapid but also remarkably resistant to inversion. It has been suggested that this sort of categorization performance could be achieved using the global distribution of orientations within the image, which interestingly is unchanged by inversion. Here, we presented subjects with two natural scenes at 16 different orientations that were simultaneously flashed in the left and right hemifield and we asked them to make a saccade to the side containing an animal. We report that human performance is surprisingly rotation invariant as reaction times were similar and accuracy remarkably stable across orientations. The results imply that this form of rapid object detection could *not* depend on the global distribution of orientations within the image. One alternative is that subjects are instead using local combinations of features that are diagnostic for the presence of an animal.

Keywords: rapid visual processing, rotation, computational models

Introduction

Performance in visual categorization tasks has been known for some time to be astonishingly fast: Humans are able to respond when a briefly flashed natural image contains an animal in as little as 250–280 ms. Parallel electrophysiological recordings show clear decision-related responses from around 150 ms (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996). These kinds of temporal constraints on visual processing have forced a rethink of traditional views of neural processing because common rate-based models cannot explain this extreme speed (Thorpe & Imbert, 1989). They also reinforce the idea that object recognition could be achieved in a feed-forward way, with a wave of spikes that passes through a hierarchy of areas of increasing complexity, before activating neurons in regions such as the primate inferotemporal cortex that can be selective to particular visual forms (VanRullen & Thorpe, 2002).

The behavioral measures supporting this view invariably include not only the time needed for visual processing but also that required for response execution, making it difficult to assess the relative contribution of each to the observed reaction times (DiCarlo & Maunsell, 2005; Johnson & Olshausen, 2003; VanRullen & Thorpe, 2001). However, in the light of recent results, it appears that processing times can be even shorter than previously thought: In a forced choice task, where two images are simultaneously flashed to the left and right of fixation, reliable saccadic eye movement responses to the side of the animal can be initiated as early as 130 ms after stimulus onset (Kirchner & Thorpe, 2006). Given that this time includes saccade preparation, this seems to imply that the underlying visual processing can be done in 100 ms or less. This seems extremely short given the values usually proposed for higher level responses in humans. For example, face selective ERPs such as the N170 typically start substantially later (Itier & Taylor, 2004; Liu, Harris, & Kanwisher, 2002; Rousselet, Mace, & Fabre-Thorpe, 2004). This raises the possibility that this sort of task is not truly making use of the highest levels of the visual system, but could rather be based on simpler heuristics that do not specifically involve the detection of animals as such, but rather low-level image attributes that happen to be associated with images containing animals. If this sort of rapid vision really depends on low-level processing only, then processing times could be expected to be very short. As shown previously, natural scene categorization could indeed be at least partially explained by an analysis of low-level features, such as can be performed in early visual areas. Specifically, Torralba and Oliva have recently proposed that the distribution of orientations, computed at spatially defined locations of the image, can be diagnostic of particular categories of images (Torralba & Oliva, 2003). More specifically, these authors have shown that a linear classifier, analyzing the distributions of energy for orientation and spatial frequency-tuned channels in a 4×4 grid, can reach accuracy levels of 80% or more on tasks such as judging whether the target category is contained in a natural image.

doi: 10.1167/6.10.1 Received October 24, 2005; published September 8, 2006 ISSN 1534-7362

Similar findings have also been reported by a different team (Mermillod, Guyader, & Chauvin, 2005) that reinforces the low-level approach to natural scene categorization.

Interestingly, in one previous study from our laboratory, Rousselet, Mace, and Fabre-Thorpe (2003) found that inverting the photographs has remarkably little effect on go/no-go performance either in the animal categorization task, or in a task requiring subjects to report the presence of animal and human faces within the scene. However, because spatially defined distributions of orientations (horizontal, vertical, and oblique) would be relatively unchanged by inversion, such a result could actually be predicted by a mechanism of the sort proposed by Torralba and Oliva (2003).

The low-level model suggested by Torralba and Oliva (2003) is based on statistics of natural images drawn from upright presentations. It follows that if the distribution of orientations is changed, through 2D rotation for example, processing times, detection accuracy, or both should be significantly impaired. In fact, the predicted decrease in performance could be directly related to the effects of mental rotation, a time-consuming operation performed by the brain to match a retinal input to internal, previously stored representations (Jolicoeur 1985; Shepard & Cooper, 1982; Shepard & Metzler, 1971; Tarr & Pinker, 1989). Experimental evidence has shown that when the observer's viewpoint changes from that used when a subject first learns to recognize an object, performance levels decrease (Bulthoff & Edelman, 1992; Christou, Tjan, & Bulthoff, 2003; Logothetis & Pauls, 1995). Considering this convergence of evidence, one would predict that if ultrarapid visual processing was based on the global distribution of energy in orientation channels, performance should be strongly affected by stimulus rotation. It would be impossible to detect targets rapidly when presented, for example, at an angle of 90°, leading either to a collapse of performance, or to a substantial increase in reaction time to allow time for mental rotation. To our knowledge, with the exception for the limited case of 180° rotation for which virtually no effect was found in animal or face categorization (Rousselet et al., 2003), no previous study has systematically investigated 2D rotation invariance for natural images.

In the present study, we employed a choice saccade task in which human subjects were presented with briefly flashed natural scenes, at 16 different orientations spread regularly around the clock, and they had to make a saccade to the side containing an animal (Figure 1).

Methods

Subjects

Sixteen volunteers (mean age = 26 ± 3.6 years, 7 women and 9 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), and all subjects gave informed consent to participate.

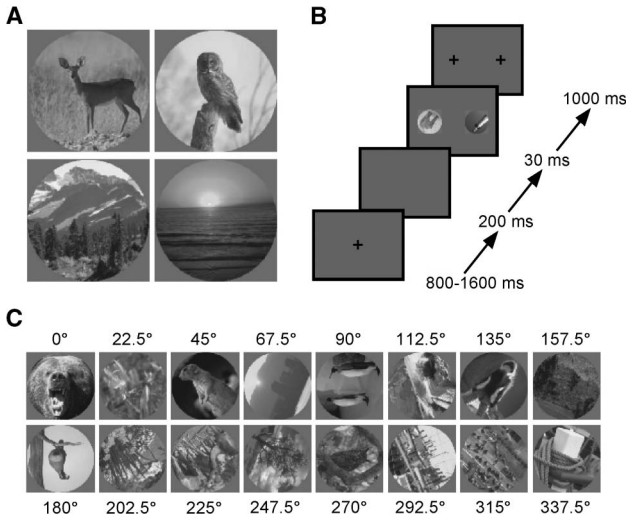

Figure 1. (A) Images were selected based on displaying an explicit straight orientation: Animal targets would thus be chosen, for example, when the legs are clearly vertical (top left) or when perched on a pole (top right); distracters where trees are vertical (bottom left) or where an horizontal was unambiguous (bottom right). (B) After a pseudorandom fixation period, a blank screen (gap period) for 200 ms preceded the simultaneous presentation of two natural scenes in the left and right hemifields (30 ms). The images were followed by two gray fixation crosses indicating the saccade landing positions. (C) Images could be rotated to 16 different angles from 0° to 337.5° by steps of 22.5° (counterclockwise). Here, examples are alternatively target and distracter images.

Experimental set-up

Subjects 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 (1,024 × 768, 100 Hz) on a background set to a median level of luminance. The monitor was at a distance of 80 cm from the subject, resulting in an image diameter of 5.6°, leaving a gap of 4.4° between the two images when presented simultaneously. The mean gray-levels of the target versus distracter images were comparable.

One thousand seven hundred commercially available photographs (768 by 512 pixels) were selected and divided in two categories: targets that included a wide range of animals in their natural environments, and distracters that depicted various landscapes without any animals present. The images were chosen because they had a natural vertical orientation: In the case of targets for example, the legs of the animal could be clearly seen to touch the ground or the animal might be sitting on a vertical pole (see Figure 1A, top); in the case of distracters, the presence of trees or an horizon line, for example, would similarly qualify them for inclusion in the stimulus set (Figure 1A, bottom).

For the purpose of the experiment, a single circular patch with a diameter of 221 pixels was manually selected and used to generate a set of 16 different images corresponding to orientations of 0° to 337.5° by steps of 22.5° (counterclockwise; Figure 1C). Among the 850 images of a given category, 50 of them were randomly chosen to be presented at all orientations to all subjects, leading to 800 presentations per subject (the "repeated" condition). For the remaining 800 images within a category, each was shown only once at a given orientation to any particular subject, again leading to a total of 800 presentations per subject (the "novels" condition).

Protocol

Two natural scenes with the same orientation were flashed for 30 ms centered at 5° in the left and right hemifield (Figure 1B). 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 black fixation cross disappeared after a pseudorandom time interval (800–1,600 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). After presentation of the images, two dark gray fixation marks were presented for 1s at ±5° to indicate the two possible saccade landing positions. The subjects performed two sessions of 10 blocks of 80 trials resulting in 50 trials per condition, per orientation and per subject (50 × 16 orientations × new vs. repeated images).

Response recording and detection

Eye position was recorded by horizontal EOG electrodes [1 kHz, low-pass 90 Hz, notch at 50 Hz, baseline correction (−400:0) ms; NuAmps, Neuroscan Inc.] 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. As a first criterion, the difference signal between the left and right EOG electrodes had to exceed an amplitude threshold of ±30 μV. Then, the saccade onset time was automatically determined as the nearest signal inflection preceding this point. Each trial was verified by the experimenter to make sure that only the largest inflection (if any) was taken as a real saccade (for more details, see Kirchner & Thorpe, 2006). A small percentage (16%; 4,014 of 25,600) of trials had to be excluded because of a noisy eye signal, but this percentage was evenly spread across conditions.

Minimum reaction times

To determine a value for the minimum SRT, we divided the saccade latency distribution of each condition into 10-ms time bins (e.g., the 120-ms bin contained latencies from 115 to 124 ms) and searched for the first bin to contain significantly more correct than erroneous responses. This allowed us to eliminate trials involving anticipations that would result in chance performance (Kalesnykas & Hallett, 1987). χ^2 tests were calculated for each condition and bin. If at least 10 subsequent tests reached significance at the $p < .05$ 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 5 correct responses.

Statistical analyses

Two-way balanced ANOVAs were systematically applied at the alpha 5% significance level, with columns standing for the orientation condition ($n = 9$) and rows for the subjects performances ($n = 16$), mean SRTs, or accuracies. The ANOVA test assumes the data to be normally distributed with equal variances. Consequently, we verified, before applying the ANOVA on any data set, that a sphericity criteria (Mauchly) for repeated measurements was fulfilled. The statistical testing of accuracy was performed on the log-transformed percentage of correct responses. This means that the reported ANOVA results [F(dl, errors) and p] for accuracy come from the transformed data. As for the clockwise/counterclockwise, symmetrical comparison and repeated/novels analyses, the same test and techniques were used, except that columns stood for the corresponding condition (clockwise, a pair of orientations or familiarity, $n = 2$).

Note that to respect the constraint of repetition in the familiarity condition, only those image pairs that had been responded to at least 14 times (out of 16 presentations) were kept for analysis. One subject having failed to reach this threshold in the repeated condition was discarded from analysis.

Results

Overall, performance measures were quite high: mean SRTs for correct trials averaged 239.8 ± 70.1 ms standard deviation (*SD*; $n = 17{,}534$; Figure 2A). The values varied considerably between subjects, from 181.1 ms (±56.9 ms *SD*; $n = 1{,}044$) for the fastest subject to 346.0 ms (±67.4 ms *SD*; $n = 654$) for the slowest one. Overall, the mean accuracy was 81.2% (17,534 target detection on 21,586 trials), although again target detection varied between subjects from 60.0% to 96.7% correct. Much of this intersubject variability can be explained by a relatively strong speed-accuracy trade-off ($r = .64$; Figure 2B). The first bin where correct responses started to significantly outnumber erroneous saccades was at 140 ms. This

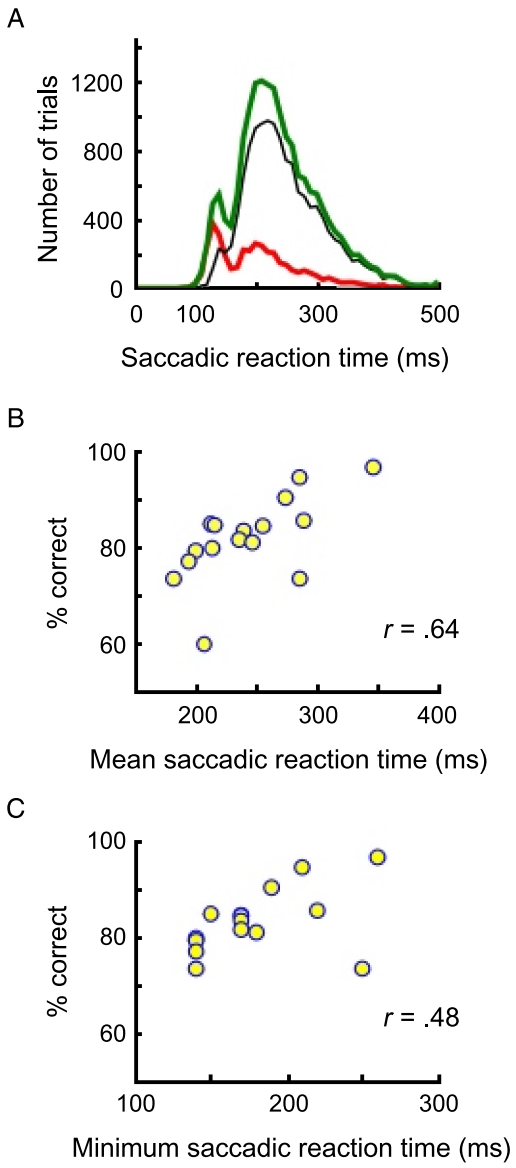

Figure 2. (A) SRT distributions of correct (green line) and incorrect (red line) responses; their difference is plotted as the black line. The minimum reaction time, defined as the first bin where correct responses started to significantly outnumber erroneous saccades, is 140 ms. First responses occurred at 120 ms with a mean reaction time of 239.8 ms. Accuracy averaged 81.2%. (B) Subject accuracy as a function of their mean reaction time. Subject performances ranged from 181.1 to 346.0 ms for accuracies between 60.0% and 96.7%, displaying a clear speed accuracy trade-off (r = .64). (C) This time, the subjects accuracy is plotted as a function of their minimum reaction time (one subject did not meet the criterion required and was discarded). Minimum reaction times were quite variable but subjects could respond as fast as 140 ms after stimulus presentation with good accuracy. Overall, again there was a speed accuracy trade-off (r = .48): the longer the subject takes to respond, the more accurate.

minimum reaction time also varied between subjects from a minimum of 140 ms to a maximum of 260 ms (Figure 2C).

Accuracy as a function of target orientation showed a remarkable level of performance: correct detection rates varied from 78.7% (worst orientation = 90°; 1,040 of 1,321) to 84.9% (best orientation = 22.5°; 1,160 of 1,366; Figure 3A). Mean reaction times on correct trials were surprisingly stable across orientations, ranging from 236.5 ms (fastest at 0°; ±69.4 ms *SD*; n = 1,125) to 244.3 ms (slowest at 270°; ±73.7 ms *SD*; n = 1,084; Figure 3B).

Pooled data

Because orientations were regularly spaced around the clock, we could first address the question of symmetry in orientation processing: Would the image be processed differently if presented in a clockwise direction (from 337.5° down to 202.5°) rather than in a counterclockwise one (from 22.5° to 157.5°)?

Overall, differences between those two conditions in terms of accuracy and mean SRTs on correct responses were extremely small and nonsignificantly different:

Accuracy: 81.9%, ±2.3 *SEM* (clockwise) versus 81.7%, ±2.2 *SEM* (counterclockwise); $F(1, 15) = 0.03$, $p > .86$.

Mean SRTs: 242.4 ms, ±11.0 ms *SEM* (clockwise) versus 242.7 ms, ±11.1 ms *SEM* (counterclockwise); $F(1, 15) = 0.14$, $p > .72$.

When comparing each orientation with its symmetric counterpart (e.g., 22.5° vs. 337.5°), again no significant differences could be detected at the 5% level either for accuracy or mean reaction times.

Because performance was equivalent for clockwise and counterclockwise rotations, we pooled the data for analysis of the rotation effects: Results for each counterclockwise orientation were grouped with those of the symmetrical, clockwise counterpart (e.g., the 22.5° and 337.5° orientations become the 22.5° one). Because upright (0°) and inverted (180°) conditions were not pooled together, we end up with 9 orientations, with upright and inverted having half as much data as the other seven.

Effect of rotation on detection accuracy

Correct detection rates varied by no more than 5.3%, with the worst mean value appearing at 135° (79.8%, ±2.3% *SEM*) and the best at 22.5° (85.1%, ±2.1% *SEM*). Although small, these variations could reach statistical significance: The percentage of correct detections displayed a statistically significant effect of orientation, $F(8, 120) = 3.68$, $p < .001$. Post hoc analyses showed that this effect was mostly due to the 22.5° orientation, which resulted in significantly better accuracy than the worst orientation at 135° (Figure 4A).

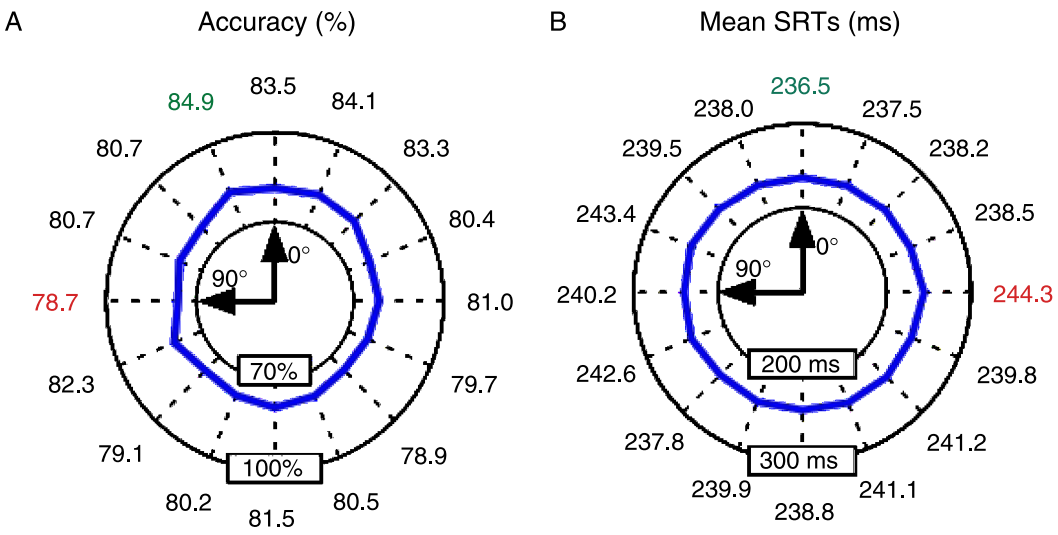

Figure 3. (A) Polar plot of the percentage of correct responses as a function of the corresponding target orientation. The inner circle corresponds to 70% of correct detections; the outer one to 100% (actual performance as the blue curve; precise values outside the polar plot). Accuracy looks remarkably stable, as it stays above 75% for any orientation (78.7% for the worst orientation at 90°, in red, to 84.9% for the best at 22.5°, in green). (B) Polar plot of mean SRTs (SRT) as a function of the corresponding target orientation. The inner circle corresponds to 200 ms; the outer one to 300 ms (actual performance as the thick curve; precise values outside the polar plot). Mean SRTs stayed within 10 ms one from another, from the fastest (236.5 ms at 0°; green) to the slowest (244.3 ms at 270°; red).

The accuracy analysis reveals a single very weak effect that affects inverted oblique presentations compared to orientations close to the vertical.

Stability of responses times

Mean SRTs on correct trials ranged from 238.8 ms (±12.4 ms *SEM*; fastest at 0°) to 244.4 ms (±11.1 ms *SEM*; slowest at 90°; Figure 4B); analysis on mean SRTs revealed no statistically significant differences between orientations, $F(8, 120) = 1.3$, $p > .25$. Reaction times for *incorrect* trials (saccades in the direction of the distracter) also failed to show any significant variations as a function of orientation, $F(8, 120) = 0.99$, $p > .45$; one subject had to be discarded in that case because he had no incorrect trials at one orientation). And finally, minimum reaction times were also quite stable because all of the nine orientations were processed as fast as 150 ± 20 ms, without any particular pattern being observable across orientations (Figure 4C). Thus, at least for reaction time measures, it is clear that ultrarapid visual processing is effectively rotation invariant.

Repeated versus novel images

Because half of the trials involved stimuli presented repeatedly, whereas the other half were only ever seen once, we were able to test whether repeated presentation improved performance. Subjects tended to respond as fast to repeated targets as to novel ones and there was no significant difference in accuracy, in line with our previous study (Kirchner & Thorpe, 2006):

Accuracy: 81.2%, ±0.9 *SEM* (repeated) versus 81.3%, ±0.9 *SEM* (novel); $F(1, 14) = 0.05$, $p > .82$;

Mean SRTs: 234.2 ms, ±3.1 ms *SEM* (repeated) versus 235.9 ms, ±3.3 ms *SEM* (novel); $F(1, 14) = 1.48$, $p > .24$.

Image orientation did not affect processing times, $F(8, 112) = 1.12$, $p > .35$, but significant effects could be observed on accuracy, $F(8, 112) = 3.04$, $p < .004$. In both conditions, 22.5° presentations stayed as best orientations and 135° as worst ones (repeated: 85.4% vs. 78.6%; novels: 86.4% vs. 79.6%). But interestingly, statistically significant effects were only observed in the "repeated" condition [repeated: $F(8, 112) = 2.19$, $p < .035$; novels: $F(8, 112) = 1.49$, $p > .17$]. In sum, the small but significant effects on accuracy we observed in the overall data set mainly occurred in the trials where subjects could see the same target at all orientations.

Discussion

Our results show that subjects can respond very quickly (mean SRT = 239.8 ms) and accurately (81.2%) when asked to detect an animal in a natural scene. Overall, accuracy was somewhat lower than in some of the previous studies from our group using paired presentations to the left and right of fixation. For example, in a go/no-go paradigm used by Rousselet, Fabre-Thorpe, and Thorpe (2002), subjects achieved 86.7% correct. Similarly, Kirchner and Thorpe (2006) reported accuracy levels of 90.1% using a

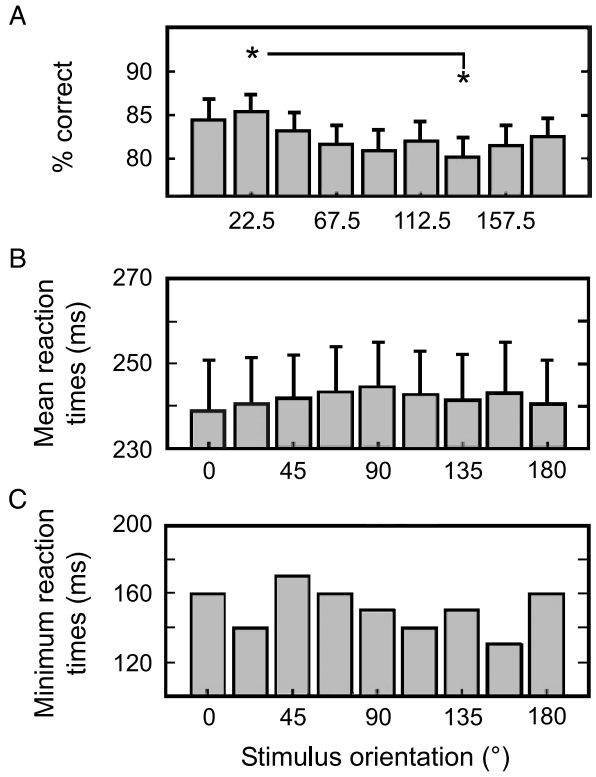

Figure 4. Mean accuracy across subjects as bars, *SEM* as error bars. Data related to a clockwise orientation (22.5° to 157.5°) could be pooled with its counterclockwise counterpart (337.5° down to 202.5°; see text for details). (A) Accuracy remains high whatever the rotation of the image, ranging from 79.8% at 135° to 85.1% at 22.5°. A statistically significant effect is detected for 22.5° compared to 135° (inverted oblique) presentations. (B) Mean SRTs display rotation invariance as they reveal no statistical effect, and they stay within 10 ms of one another, from the fastest orientation (238.8 ms at upright) to the slowest one (244.4 ms at 90°). (C) As in the general case, minimum reaction times are quite stable, as they range between 130 and 170 ms. They also display rotation invariance.

choice saccade task that was very similar to the one used here. However, there are other differences between the experiments, apart from the use of rotated images that could well explain the somewhat lower levels of performance seen here. One of the most important is the fact that the images in the Kirchner and Thorpe study were shown at lower resolution resulting in a size of 10° whereas we showed them with a size of 5.6°. Although the eccentricities of the images were about the same (here 5° rather than 6° in the previous study), the spatial gap between the images was correspondingly larger (4.4° in the present experiment, compared with 2° in the former). This means that the subjects are forced to use more peripheral parts of the visual field, and this could well contribute to the somewhat lower levels of performance seen here.

Nevertheless, the present findings illustrate the already well-documented speed of the ventral visual system (Keysers, Xiao, Földiák, & Perrett, 2001; Kirchner & Thorpe, 2006; Rousselet et al., 2002; Thorpe & Fabre-Thorpe, 2001; VanRullen & Thorpe, 2001; Wallis & Rolls, 1997). Although it has already been reported that the ability of humans to detect faces and animals in natural scenes is only weakly affected by inversion (Rousselet et al., 2003), the present results extend these observations to 2D target rotation at a large number of orientations. In fact, reaction times were remarkably stable across all the 16 regularly spaced angles, and accuracy was always above 75% and slightly, but constantly, better when presentations were made at upright or near-upright angles. It is thus clear that 2D rotation of the stimuli produces only minor effects on the efficiency of ultrarapid animal detection in natural images. As the orientation angle increasingly differed from the (near)upright position, we observed a mild drop in performance that peaked when the images were presented between 90° and 135°. Performance then recovered slightly to reach close to normal values when the image was completely upside down. Here, we can note that the orientations close to 90° where the reaction times appear to be longest are also the ones where accuracy was poor. Although weak, the fact that both effects go in the same direction suggests that there may be some effect of orientation. However, the present results demonstrate that the effects of unusual orientations are far weaker than might have been thought. Thus, although the rotation invariance of ultrarapid visual processing is very good, it is perhaps not complete.

But did subjects really look for an animal in the images or could they have simply been reacting to the presence of a foreground object, whatever its category might be? It is conceivable that they might have used a strategy of checking whether there was a large target object in the foreground. Such a strategy would indeed be relatively rotation invariant because the animal-like shape, which can be here characterized as a "blob" in the center of the image, would have a similar appearance whatever the orientation of the images. To address this issue, we performed a post hoc analysis, using only trials in which the same target was presented repeatedly. Of the 50 possible repeated stimulus pairs, 8 had a distracter in which a salient object was present in the foreground (Figure 5). We found that subjects were virtually as accurate with such pairs as when the distracter foreground contained no salient object: 79.1%, ±3.2 *SEM* (salient) versus 81.8%, ±2.2 *SEM* (not salient); $F(1, 14) = 1.81$, $p > .2$. They also responded about as fast in both conditions: 237.1 ms, ±10.8 ms *SEM* (salient) versus 234.0 ms, ±9.0 ms *SEM* (not salient); $F(1, 14) = 0.39$, $p > .5$. Although not statistically significant, the differences between these two conditions may suggest that the "not salient" one tends to give faster and more accurate responses to the target. Interestingly, this trend could be observed in the case of unrepeated presentations too ("novel" condition—see above). Here, we performed a post hoc analysis on the novel trials that used any one of the 158 distracter images identified as containing a salient

Figure 5. Eight of the 50 repeated pairs possessed a salient distracter (right column; the paired target in the left column). Each subject saw these pairs at the 16 orientations.

object in the foreground. Subjects tended to respond 8 ms faster in the "not salient" condition than in the "salient" one (234.3 ± 9.2 ms *SEM* vs. 241.9 ± 11.2 ms *SEM*; $F(1, 14) = 3.18$, $p > .07$). They were also significantly more precise [82.2% ± 2.0% *SEM* vs. 78.0% ± 2.3%; $F(1, 14) = 11.73$, $p < .005$]. Because the presence of a foreground distracter actually impairs processing, it appears that the visual system can use the presence of a foreground object as a cue for performing the task. Note, however, that the 4.2% accuracy difference between the two conditions is still relatively small, and subjects were still able to perform remarkably good in the salient condition, when both "novel" and "repeated" images were used. These results make it very unlikely that the subjects could have performed the task only by detecting the presence of a foreground object.

As a consequence, it is not completely clear what the subjects actually do when they perform the task; that is, more specifically, which level of processing is being used. One possibility is that they might use global scene properties to infer which scene is more likely to contain an animal. In this case, they might be relying mainly on their ability to categorize the scene, and there is evidence that this sort of scene categorization can be achieved based on relatively low-level information such as the distribution of energy in spatial frequency- and orientation-tuned channels (Torralba & Oliva, 2003). Alternatively, they might be relying on their ability to detect particular local feature combinations that are in some way diagnostic for animals. These might include an ability to detect specific forms, such as an eye, a fin, or a beak, which would qualify as intermediate representations. But it might also be the case that the subjects use these different levels of analysis together, and indeed the robustness of our ability to perform this sort of task suggests that there will not be a single strategy. Nevertheless, the relatively modest effects of rotating the entire image argue against the exclusive use of the simplest strategy, based on local distributions of energy at different orientations. In the paper by Kirchner and Thorpe (2006), further specific tests also allowed us to rule out the possibility that low-level clues such as contrast, that just happen to correlate with the presence of an animal, could explain the subjects ability to perform the task. As a consequence, we propose to take into account a hierarchical approach, where a lower level of processing is necessary but not sufficient, to explain the remarkable performances of the visual system reported here.

It is worth commenting here on the apparent discrepancy between our data showing a surprising degree of orientation invariance and the well-known literature on the effects of inversion on face processing. For example, the famous Thatcher illusion (Thompson, 1980) demonstrates very well that processing of inverted stimuli can be very different to processing of normally oriented stimuli. However, it is important to realize that in our task, subjects are not required to make any form of identification judgment—they simply have to respond when an animal is present.

It could be that animal detection and animal identification differ greatly, in that the former would be based on rapid feature analysis alone whereas the latter would involve attentional resources in a subsequent step (Evans & Treisman, 2005). According to this scheme, we showed here that the first, rapid feature analysis is remarkably robust to orientation. Further experiments will be needed to determine whether the ability of subjects to perform a more demanding task (e.g., reporting the species of an animal) shows the same degree of robustness (Riesenhuber & Poggio, 2000).

This could also explain why mental rotation, the typical object recognition mechanism expected to be involved in this task, can hardly be invoked in our case. Even if our experiment differs from earlier studies in that the set of stimuli to be recognized was very much larger, and that rotations were done in the picture plane only, it was nevertheless quite natural to expect substantial perturbations in performance with rotated images. Specifically, if some form of mental rotation was involved, increases in reaction time of tens or even hundreds of milliseconds could have been anticipated (Biederman & Bar, 1999; Vanrie, Willems, & Wagemans, 2001). However, the reasons underlying the increases in reaction time with nonstandard views are somewhat controversial. Whereas some authors have argued for a true "mental rotation" phenomenon, Perrett, Oram, and Ashbridge (1998) demonstrated that similar reaction time changes could be produced by a process of evidence accumulation at the level of single neurons. They noted that when a neuron is stimulated with an optimal stimulus, the firing rate builds up very rapidly, but that the slope of the firing rate curves is flatter with less optimal stimuli. Thus, when an object is rotated away from the optimal orientation, it would take longer for the neuron to reach a given threshold firing rate. In this way, it is possible to explain why reaction times could increase as the stimulus is rotated away from the standard view without having to postulate anything like the rotation of an internal model.

Interestingly, Perrett et al. (1998) also noted that the degree to which reaction time depends on orientation is highly dependent on the criterion used. If the system needs to accumulate a large number of spikes to make a decision, the time to reach the threshold will vary a great deal with orientation. In contrast, if only a few spikes are needed from each neuron, decisions can be made rapidly, and furthermore there is much less dependence on orientation. From this point of view, the current results, which provide evidence for rapid decisions that are remarkably insensitive to changes in orientation, suggest that the underlying processing uses only a very small number of spikes per neuron. This is an idea that has been previously proposed as a way to understand rapid scene processing (VanRullen & Thorpe, 2002).

Indeed, biologically inspired image processing algorithms such as SpikeNet do suggest that even quite complex visual forms can be detected rapidly based on the order of firing within adjacent columns of neurons (Thorpe, Guyonneau, Guilbaud, Allegraud, & VanRullen, 2004). This form of order-based recognition mechanism, which only uses a single spike in a relatively small percentage of cells, can operate very rapidly and has been found to show a remarkable degree of orientation invariance. For example, when a shape such as an eye or a mouth has been learned, the same recognition mechanism will respond over a range of ±10–15° of rotations in the picture plane. Clearly, a single such mechanism would be incapable of responding accurately to the range of image orientations used here. However, it may be enough to have been exposed to a wide range of diagnostic feature combinations at a wide range of angles for an experience-based system to be able to respond over a full range of orientations. There is now increasing evidence that this form of view-based recognition mechanism plays an important role in object vision. If such a view is correct, the relatively small variations in efficiency with changing orientation could simply result from the fact that we possess mechanisms sensitive to local diagnostic features, and that we have enough of them to cope with situations where the animal parts are presented at a wide range of orientations.

Acknowledgments

Research supported by the CNRS, the ACI "Neurosciences Computationelles et Intégratives," the SpikeNet Technology SARL, the and European Research Training Network "Perception for Recognition and Action" (HPRN-CT-2002-00226).

Commercial relationships: SpikeNet Technology.
Corresponding author: Rudy Guyonneau.
E-mail: rudy.guyonneau@cerco.ups-tlse.fr.
Address: Centre de Recherche Cerveau et Cognition, 133 route de Narbonne, 31062 Toulouse Cedex, France.

References

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39,* 2885–2899. [PubMed]

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89,* 60–64. [PubMed] [Article]

Christou, C. G., Tjan, B. S., & Bulthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel

viewpoints. *Journal of Vision, 3*(3), 183–198, http://journalofvision.org/3/3/1/, doi:10.1167/3.3.1. [PubMed] [Article]

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

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492. [PubMed]

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,* 171–180. [PubMed]

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16,* 553–567. [Article]

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14,* 132–142. [PubMed] [Article]

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3*(7), 499–512, http://journalofvision.org/3/7/4/, doi:10.1167/3.7.4. [PubMed] [Article]

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289–303. [PubMed]

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

Keysers, C., Xiao, D.-K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience, 13,* 90–101.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776. [PubMed]

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

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex, 5,* 270–288. [PubMed]

Mermillod, M., Guyader, N., & Chauvin, A. (2005). The coarse-to-fine hypothesis revisited: Evidence from neuro-computational modeling. *Brain and Cognition, 57,* 151–157. [PubMed]

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalization of recognition without mental transformations. *Cognition, 67,* 111–145. [PubMed]

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3 Supplement,* 1199–1204. [PubMed] [Article]

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

Rousselet, G. A., Mace, M. J., & 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, http://journalofvision.org/3/6/5/, doi:10.1167/3.6.5. [PubMed] [Article]

Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision, 4*(1), 13–21, http://journalofvision.org/4/1/2/, doi:10.1167/4.1.2. [PubMed] [Article]

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703. [PubMed]

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233–282. [PubMed]

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9,* 483–484. [PubMed]

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

Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist modelling. In R. P. et. al. (Ed), *Connectionism in perspective* (pp. 63–93). Elsevier.

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Neuroscience. Seeking categories in the brain. *Science, 291,* 260–263. [PubMed]

Thorpe, S. J., Guyonneau, R., Guilbaud, N., Allegraud, J.-M., & VanRullen, R. (2004). SpikeNet: Real-time visual processing with one spike per neuron. *Neurocomputing, 58,* 857–864.

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

Vanrie, J., Willems, B., & Wagemans, J. (2001). Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features. *Perception, 30,* 1047–1056. [PubMed]

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

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

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51,* 167–194. [PubMed]