Faces in the cloud: Fourier power spectrum biases ultrarapid face detection

Christian Honey
MSc Program in Neuroscience, Oxford University, Oxford, UK
Université de Toulouse, CerCo, UPS, Toulouse, France, & CNRS, UMR5549, Faculté de Médecine de Rangueil, Toulouse, France

Holle Kirchner
Université de Toulouse, CerCo, UPS, Toulouse, France, & CNRS, UMR5549, Faculté de Médecine de Rangueil, Toulouse, France

Rufin VanRullen

Recent results show that humans can respond with a saccadic eye movement toward faces much faster and with less error than toward other objects. What feature information does your visual cortex need to distinguish between different objects so rapidly? In a first step, we replicated the "fast saccadic bias" toward faces. We simultaneously presented one vehicle and one face image with different contrasts and asked our subjects to saccade as fast as possible to the image with higher contrast. This was considerably easier when the target was the face. In a second step, we scrambled both images to the same extent. For one subject group, we scrambled the orientations of wavelet components (local orientations) while preserving their location. This manipulation completely abolished the face bias for the fastest saccades. For a second group, we scrambled the phases (i.e., the location) of Fourier components while preserving their orientation (i.e., the 2-D amplitude spectrum). Even when no face was visible (100% scrambling), the fastest saccades were still strongly biased toward the scrambled face image! These results suggest that the ability to rapidly saccade to faces in natural scenes depends, at least in part, on low-level information contained in the Fourier 2-D amplitude spectrum.

Keywords: face processing, natural scenes, image scrambling, Fourier transform, wavelet transform, power spectrum, phase scrambling


Introduction

A happy guy stands next to his new sports car. What information of that scene lets us recognize, seemingly at a glance, that there is a face in the scene—or a car?

Recent results from our lab suggest that we can categorize different scenes (faces/means of transport/animals) with high accuracy (~90%) and an astonishing speed of 120 ms with saccadic responses (Kirchner & Thorpe, 2006). This brings forth the question if, in such short time, retinal information is processed “up” to a semantic level, where the identity of an object may be represented explicitly (Reddy & Kanwisher, 2006).

Faces enjoy an advantage over other scene categories: In a recent eye tracking study, Fletcher-Watson, Findlay, Leekam, and Benson (2008) found that subjects had a strong bias for looking toward images containing faces when presented side by side with images without faces. This is compatible with evidence from our group, demonstrating that very early and selective saccades (even before 120 ms) can be directed toward faces but not toward other object categories (Crouzet, Thorpe, & Kirchner, 2007). Some investigators report that faces pop-out of a scene, though the authors are still debating whether the pop-out effect depends on semantic, high-level information of faces (Hershler & Hochstein, 2005), or bear on low-level information such as the Fourier amplitude spectrum (VanRullen, 2006). We will here report evidence in favor of the latter perspective.

Classical models of overt visual attention (Itti & Koch, 2000; Treisman & Gelade, 1980) assume that our first saccades are driven by low-level features such as luminance, orientation or color contrast. Only after building a first representation of their respective saliencies will attention play a role in semantic scene processing. Thus, in this perspective, object categorization would be performed only after at least one fixation, allowing for accumulation of sufficiently detailed information of a region in visual space.

A number of studies have investigated the causal role of low-level features in the selection of fixation locations (Einhäuser & König, 2003; Parkhurst, Law, & Niebur, 2002) with the one common finding that fixated scene
locations contain higher luminance contrast than random control locations. But the fact that object boundaries are generally defined by foreground/background luminance contrast differences leaves open the alternative interpretation: that objects are identified before and drive early saccades. This alternative is supported by the results on ultrarapid categorization cited above, and by a recent study showing that in large scenes containing a face among other objects, the very first saccades tend to be directed to the face (Cerf, Harel, Einhaeuser, & Koch, 2008).

To summarize, the reported work suggests that objects can be visually categorized with remarkable speed, in the

Figure 1. Examples of (A) face images and (B) transport images in both scrambling conditions: top rows: wavelet scrambling; bottom rows: phase scrambling. Scrambling levels from left to right: 0%, 30.25%, 55%, and 100% (see Methods). (C) Example images from our stimulus set; top row: faces; bottom row: means of transport.
range of around 120 ms after stimulus onset. Further, humans show a strong bias toward fixating faces, when given a choice between different object categories. Here we ask what part of the visual information content about faces renders them favorable to the human visual system. Are explicit face representations activated so early? Or could low-level information suffice to elicit this early face preference?

We conducted a series of visual discrimination experiments to test if a particular aspect of the physical information in face images can drive their apparent rapid categorization. We presented to our subjects two images simultaneously on a computer monitor, one containing a face, the other one a means of transport. Each image had one of five contrast levels and we asked our subjects to saccade to the image with higher contrast.

This task allowed us to measure biases toward either category, irrespective of individual categorical preference and related search strategies: The semantic category of the image was entirely irrelevant to the task.

To assess the influence of particular low-level image properties, we presented face/transport pairs at 4 different levels of “image scrambling.” One group of subjects saw image pairs that were scrambled in the Fourier phase domain, keeping the orientation of all spatial frequency components (i.e., the 2-D amplitude spectrum) constant. The other group saw wavelet-scrambled images, changing the orientation of local spatial frequency components while keeping the wavelets position (and strength) constant (see Figure 1).

Thus we manipulated either the position or the orientation of the spatial frequency content of the image pairs. If the bias in rapid face categorization is driven by either of these features (orientation or position of spatial frequency components), a categorization bias toward face images should be preserved in fully scrambled images that kept the critical feature content constant. Is rapid scene categorization semantic or low-level?

Our results indeed suggest that the early face preference partly relies on the plane orientation information of the 2-D power spectrum of face images, without the need for the activation of explicit representations.

Methods

Subjects

All subjects were between 18 and 32 years old and had normal or corrected-to-normal visual acuity. They all gave informed consent and were naive to the purpose of the experiment. None of them had seen the stimuli before. Initially, 19 subjects were tested; however, as reported in more detail in the Results section, three of these did not display the expected face bias and were thus not included in subsequent analysis. Among the remaining 16, one group of 8 subjects was assigned to the “wavelet scrambling” condition, the other group of 8 subjects performed the “Fourier phase scrambling condition.”

Protocol

We presented to our subjects two stimuli at the same time on a computer CRT monitor (Sony Trinitron Multiscan G400) in a darkened room. They held their heads stable with a chin rest, such that their eyes were positioned centrally 55 cm away from the screen. The presented grayscale images all were 300 × 300 pixel in size, which made ∼10° visual angle, with their inner borders 2° and the image centers 7° horizontally away from the screen center.

One trial consisted of the following steps:

1. a central fixation cross appears on the screen for 200 ms,
2. the fixation cross disappears for 100 ms,
3. the two stimuli appear for 400 ms, and
4. the stimuli disappear, blank screen for 800–1600 ms.

In the main experiments, our subjects’ task was to “look as fast as possible at the image with higher contrast.” We told subjects that it would suffice to look at the center of the chosen image and return gaze to the screen center when the fixation cross re-appeared.

We measured the subjects’ eye movements with an ISCAN ETL 200 eye tracker. The subjects sat down on a chair at the table with the presentation monitor on top. We asked our subjects to hold their head still during stimulus presentation. After each block of 108 trials the subjects could move freely if and for as long as they wanted.

The eye scanner uses two infrared cameras monitoring pupil center position and corneal reflection position through semi-transparent mirrors. An infrared light bulb is mounted next to each camera, providing the light reflected from the eyes.

Before and after each block, we performed a nine-point calibration on all screen coordinates ±15°, 0° horizontally, and ±10°, 0° vertically from the screen center. From the measured angles, we estimated the actual fixation angles for saccade analysis.

Stimuli

We used a total of 300 original gray-scale images, 150 faces and 150 means of transport. All images were normalized for mean luminance and contrast (defined as the standard deviation of the luminance distribution of an image, divided by that image’s mean luminance). For each
subject, we randomly generated 150 image pairs, with one face image and one transport image in each pair, examples of which are shown in Figure 1C.

For each image pair, we then generated 4 scrambling levels, according to two different scrambling methods (see below): both images in a pair were scrambled to the same extent with scrambling levels 0%, 30.25%, 55%, and 100% (the last 3 levels providing a linear progression on a log scale). The database thus constituted included 150 × 4 levels = 600 image pairs for each scrambling method.

The subjects’ task was to gaze, as quickly as possible, to the higher contrasted image. Therefore we had to change for each image pair the images contrast. Initially we equalized for both mean luminance and standard deviation the whole database of 600 face and 600 transport images. Then we defined 5 contrast levels: 16%, 25.3%, 40%, 63.25%, and 100% (providing a linear progression on a log scale). For a given image pair, we assigned to the face image the medium contrast level 3 (40%) and to the transport image one of the 5 possible values. We also generated the inverse condition, where for the same image pair, the transport image had contrast level 3 and the face had a variable contrast value. This yielded a total of 1200 image pairs. In those conditions where both images had contrast level 3, we removed the inverse condition (which was identical) to avoid duplicating a particular trial. This resulted in 1080 image pairs. We distributed these 1080 image pairs randomly in 10 presentation blocks such that each block contained 108 trials (i.e., image pairs), ensuring that inverse conditions (for one image pair) were not presented within the same block.

**Image scrambling methods**

We wanted to test if selected low-level features could yield a bias toward faces even in the absence of semantic face information. Therefore we presented to the first subject group image pairs scrambled in the wavelet domain, keeping the positions (and coefficient strengths) but randomly changing the orientations of the wavelets. The other group saw image pairs that were scrambled in the Fourier domain, keeping the orientation, but randomly changing the phase of spatial frequency components (Figures 1A and 1B).

**Wavelet scrambling**

For wavelet scrambling, we used the 2-D discrete wavelet transform with a Dechaubie-7/9 mother wavelet. The transform yielded the wavelet coefficients on 8 spatial scales for the horizontal, diagonal (45° from top left), and vertical wavelet content of a given original image. On all 8 scales of the wavelet transform, we randomly switched the orientations of a certain fraction of the wavelet coefficients for that scale: for example, 55% scrambling means that at 55% of the positions for a given scale, the three orientation wavelet coefficients were randomly permuted (the coefficient for the vertical orientation taking the value of the horizontal or diagonal coefficient, etc.). The wavelet scrambling thus changed the orientation at affected positions (thus removing contours and local structure, which carry semantic and object information) but maintained the wavelet coefficient values at each position (thus maintaining the spatial distribution of energy in the image).

**Phase scrambling**

For phase scrambling, we used the 2-D Fast Fourier Transform. In the Fourier representation of each image, we randomized a certain fraction of all Fourier coefficients: for example, 30.25% scrambling means that 30.25% of all Fourier coefficients were shifted to a new, random phase (by adding a random value between +π and −π). Note that this scrambling method does not suffer from the potential confounds described by Dakin and collaborators, which occur when phase information from an image is mixed with phase information from a noise pattern (Dakin, Hess, Ledgeway, & Achtman, 2002; Rainer, Augath, Trinath, & Logothetis, 2001).

The phase scrambling method changed the position (i.e., the phase) of all affected spatial frequency components but maintained the Fourier 2-D amplitude spectrum across orientations and spatial frequencies. Similar to the wavelet scrambling, it thus allowed us to disrupt the semantic content of each image. However, here the distribution of energy across the orientations (i.e., the 2-D power spectrum) was maintained while their positions (the phase spectrum) were randomized, whereas the wavelet scrambling provided the opposite manipulation (randomizing orientation while retaining position information).

**Control task**

With both methods, we generated images of 0%, 30.25%, 55%, and 100% scrambling. In an additional judging task, we ensured that we had successfully removed all semantic information from the image at the 100% scrambling level: A separate group of 4 subjects saw the 100% scrambled images, one image at a time, for as long as they needed. We asked them to indicate with a mouse button press whether or not they could identify an object (of any category) in the image. If they reported seeing an object they could additionally choose, with another mouse button press, what kind of object it was. Eight choices were provided: [cars, faces, forest, mountains, cities, indoors, clouds]. Only two out of the 300 scrambled face images (150 phase scrambled + 150
wavelet scrambled) were correctly categorized as faces by two or more subjects. Not a single transport image was correctly categorized by more than one subject.

### Data analysis

#### Definition of saccades

We defined the direction of a saccade after stimulus onset as the side where the measured horizontal visual angle first crossed one of the images inner borders, 2° from the screen center. The onset of that saccade was then defined as the sample point after stimulus onset when the derivative of the measured angle last changed sign and maintained the sign up to the moment when the angle crossed the image border indicating that the eye movement toward the measured direction was consistent before hitting the image (adapted from Kirchner & Thorpe, 2006).

#### Performance measure

In every trial, one of the two images had a contrast value of 40% (the reference) and the other (the probe) a randomly determined level between 16% and 100%. We measured probe choice probability as a function of probe contrast, separately for faces and means of transport as the probe category. We then fitted the two resulting curves with cumulative normal functions (using the “CumNormYNFitFun” function of the Matlab Psychophysics Toolbox).

#### Bias measure

If subjects systematically tend to prefer faces over means of transport (as previous studies have suggested), they should saccade to face images more often than to transport images when both images have the same contrast level 3 (and no correct decision is possible). Then, equal

Figure 2. Probe choices in both subject groups (A: wavelet group; B: phase group) for the unscrambled images as a function of the contrast of the probe (triangles represent the choice of faces and squares the choice of transport). The non-probe (reference) stimulus always had a contrast of 0.4. Error bars represent standard error of the mean.
probability of choosing either category should be reached at a lower contrast than level 3 when the face is the probe, and at a higher contrast when the means of transport is the probe. We thus defined “face bias” as the horizontal distance between the two psychometric curves at 50% choice level: The amount by which face contrast can be lowered, plus the amount by which transport contrast has to be increased, to yield chance performance in the contrast discrimination task.

**Biases at different scrambling levels and with different scrambling types**

We checked the bias strength for significant differences from zero at all 4 scrambling levels with Student’s t-test. Then we split each subjects’ data into “fast” and “slow” saccades, according to the median reaction time from this subjects’ reaction time distribution. We analyzed face biases of all subjects for fast and slow saccades and compared the results over all 4 scrambling levels and over scrambling type with a balanced, repeated measures 3-way ANOVA (with bias strength as the independent variable).

**Results**

Our first step was to investigate whether a bias to saccade faster and/or more often toward face images (as observed in several studies mentioned in the Introduction) could be replicated within our experimental paradigm. For this, we consider only the original, non-scrambled image pairs.

In each image pair, either the face or the transport image had a luminance contrast of 40%. This image we refer to as the reference stimulus. The second image of the pair had one of 5 possible contrasts and is referred to as the probe stimulus. Figures 2A and 2B show for both subject groups the respective psychometric curves for probe choice probability as a function of probe contrast (separately for face and transport as the probe) in the contrast discrimination task. If subjects tend to saccade more often toward faces, then a lower face contrast might be needed to reach an equal proportion of face vs. means of transport choices when the means of transport image is at the reference contrast. Similarly, a higher contrast would be needed for the means of transport image when the face image is at the reference contrast. Therefore, the “face bias” was defined as the contrast distance between 50% face choice (with face as the probe) and 50% transport choice (transport probe), as shown in Figure 2.

When faces were the probe stimuli (transport contrast constant at 0.4) it took a large decrease (>0.15 for both groups) in face contrast to get to an equal (50%) face choice. Inversely, when the probe was a transport image (face contrast constant at 0.4), it required a large increase (>0.25 for both groups) of contrast to make subjects choose transports over faces. Note that at 0.4 contrast

![Figure 3](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933525/) Mean biases toward face images at 5 different levels of scrambling. Circles: phase scrambled images; squares: wavelet scrambled images. The face bias in both groups is not significantly different at any level of scrambling. Error bars represent 95% confidence intervals.
Figure 4. Face biases in both groups (A: phase group; B: wavelet group) for fast and slow saccades. The means of the median reaction times in each group are indicated in each plot. At 100% scrambling the biases in the wavelet group are no longer significant. In contrast the bias for fast saccades toward phase-scrambled faces remains very strong and highly significant. Error bars represent 95% confidence intervals.
level both images had the same contrast, so no correct choice existed. Yet in this case subjects systematically (i.e., more than 70% of the time) saccaded toward the face.

We observed this “face bias” for 16 out of the initial pool of 19 subjects. Interestingly one of the three subjects who displayed the opposite “car bias” reported fear of cars, as he had been involved, as a biker, in two car accidents before. Thus, the face bias might be subject to learning or conditioning (the other two subjects did not report anything noteworthy).

In summary, we could replicate a significant bias of human observers to direct their first saccades toward faces (chi-square test for 16 occurrences out of 19 subjects, $\chi^2 = 8.9$, $df = 1, p < 0.005$). For further analysis, we focused on the 16 subjects who showed a positive face bias (these 16 subjects constitute the 2 groups of 8 subjects whose results are reported in Figures 2-4).

Next we checked if our subjects maintained the face bias when progressively scrambling both presented images, preserving selected low-level information. For the “wavelet group” we kept the local energy content constant while randomly changing local wavelet orientations on all spatial scales (wavelet scrambling), for the “phase group” we kept the orientations at all spatial frequencies constant while randomizing phase. The average biases for both groups are shown in Figure 3.

With increasing scrambling the face bias decreased for both groups. But at 100% scrambling our subjects still showed a small (phase: 0.1008; wavelet: 0.1068) and significant tendency to prefer faces over means of trans-saccadic responses and was much smaller (and in fact, not significant) than the one obtained with phase scrambling method: Disrupting the phase of Fourier components (i.e., their relative positions, but not their orientation) did not abolish the face bias. The disruption of the orientation content of the scene with our wavelet scrambling method also left a small amount of bias toward faces; but this bias did not depend on the speed of saccadic responses and was much smaller (and in fact, not significant) than the one obtained with phase scrambling when only the fast responses were considered. We therefore conclude that the 2-D amplitude spectrum of faces across orientations contains a major driving factor for fast face biases in object categorization.

As a control, we compared the face biases in both groups, only for unscrambled images (leftmost data points in Figures 4A and 4B). A balanced 2-way ANOVA showed no main effect of scrambling type but an effect of reaction time ($F(1,31) = 5.19, p < 0.05$). This comparison did not show interaction between scrambling type and reaction times. Thus the face bias can be considered similar in both groups for those images that are not scrambled. This was highly expected since, without scrambling, the images were exactly the same for both the phase and wavelet groups. As a second control, we compared the results for 100% scrambled images between both groups. A significant interaction ($F(1,31) = 7.29, p < 0.05$) between reaction time and scrambling type was confirmed, showing that early responses are mostly affected by phase information.

**Discussion**

Our results suggest that the fast face bias partly relies on low-level scene information, in particular the 2-D amplitude spectrum across orientations, as preserved with our phase scrambling method: Disrupting the phase of Fourier components (i.e., their relative positions, but not their orientation) did not abolish the face bias. The disruption of the orientation content of the scene with our wavelet scrambling method also left a small amount of bias toward faces; but this bias did not depend on the speed of saccadic responses and was much smaller (and in fact, not significant) than the one obtained with phase scrambling when only the fast responses were considered. We therefore conclude that the 2-D amplitude spectrum of faces across orientations contains a major driving factor for fast face biases in object categorization.

**Is a particular part of the amplitude spectrum responsible for the face bias?**

Whether face detection or recognition builds on a limited spatial frequency band of the amplitude spectrum is as yet an unsettled issue. Studies on human face recognition in the last 25 years, using variable methods of spatial frequency filtering, yielded contradictory evidence as to what band of spatial frequencies is sufficient and optimal for face recognition (Costen, Parker, & Craw, 1994, 1996; Fiorentini, Maffei, & Sandini, 1983; Gold, Bennett, & Sekuler, 1999; Näsänen, 1999).

One hypothesis, postulating coarse-to-fine usage of spatial frequency in face processing, relies first on physiological evidence for faster processing along the magno-cellular pathway (tuned to lower spatial frequencies; Nowak & Bullier, 1998), and second on the observation of stronger occipital ERP activations to LSF...
face content (as compared to HSF content) in face detection experiments (Holmes, Winston, & Eimer, 2005; Pourtois, Dan, Grandjean, Sander, & Vuilleumier, 2005). In contrast, a recent study on face detection (Halit, de Haan, Schyns, & Johnson, 2006) showed that both the face-sensitive N170 ERP-component as well as performance of detection suffered upon removing high spatial frequencies (>24 cycles/image) from faces. Their results confirm those of Fiorentini et al. (1983) and support the hypothesis that the visual system flexibly exploits those spatial frequencies that are relevant to the task at hand (Flevaris, Robertson, & Bentin, 2008; Morrison & Schyns, 2001; Schyns & Oliva, 1997, 1999).

In summary these findings indicate that for face detection a wide range of spatial frequencies might be necessary for optimal performance.

In addition, it must be emphasized that the critical information for producing fast face biases is not just contained in the 1-D amplitude spectrum (collapsed over orientations) but that it is instead a property of the full 2-D amplitude spectrum across the different orientations. Indeed, numerous previous studies have revealed that the 1-D amplitude spectrum, and in particular its power coefficient $a$ (verifying $A(f) \propto 1/f^a$, where $A$ denotes the amplitude and $f$ the spatial frequency) can differ systematically across different image categories (Parraga, Troschianko, & Tolhurst, 1992). In our case, the face images in our database generally had a higher $a$ coefficient than the transport images (Figures 5A and 5B). However, this was true at all levels of the phase and wavelet scrambling conditions. In fact, when an “ideal” observer was simulated, who systematically saccaded to the image with higher $a$ coefficient, we found that this observer would have chosen the face image on about 80% of trials in all conditions (Figure 5C). While this could possibly account for the choices of our human observers when no scrambling was applied to the images (about 75% of face choices in this case, assuming equal contrast between the face and transport images), it would fail to explain the behavior of these subjects when scrambling was increased (Figure 5C) as well as the interaction between scrambling method (phase/wavelet) and reaction time (fast/slow) observed in Figure 4.

Figure 5. Comparison of the 1-D amplitude spectra between face and transport images in the different scrambling conditions. Average $a$ coefficients of the amplitude spectrum (verifying $A(f) \propto 1/f^a$) are shown for the face images (in red) and the transport images (in blue) across the different scrambling levels: (A) wavelet scrambling, (B) phase scrambling. Error bars represent standard deviation. Face images have consistently higher amplitude coefficients than transport images; however, this remains true irrespective of scrambling level or condition. (C) An “ideal” observer consistently choosing the image with the highest amplitude coefficient would saccade to the face in about 80% of trials for all scrambling levels and conditions. For comparison, the proportion of face choices for our human observers at equal contrast between the face and transport image (i.e., 40% contrast) is reported on the same graph (filled areas represent standard error of the mean).
In conclusion, the fast face bias depends on the 2-D distribution of energy across spatial frequencies and orientations rather than the 1-D distribution of energy across spatial frequencies—commonly coined the “amplitude spectrum”; in addition, the comparison between phase and wavelet scrambling methods reveals that the fast face bias does not depend much on the absolute or relative spatial distribution of energy in the images, i.e., on phase information.

What brain mechanism could be responsible for this low-level processing of a semantic category?

The fusiform face area is known to show a significantly stronger BOLD signal to faces as compared to other objects (Kanwisher, McDermott, & Chun, 1997; Sergent, Ohta, & MacDonald, 1992).

Yue, Tjan, and Biederman (2006) asked “what makes faces special?.” Their subjects determined whether 2 successively presented images showed identical objects while measuring their FFA and LOC activity using fMRI. The presented image pairs were filtered in the Fourier domain to contain either the same frequency/orientation content or complementary parts of that information (i.e., changing the combination of spatial frequencies and their orientation). First, when successively presenting two face images with complementary frequency/orientation information they observed a significant drop in identification performance. In correct recognition trials they observed a correlated release-from-adaptation. In contrast, with control blob stimuli they did not observe these effects, neither in FFA nor in LOC. They concluded that “what makes the recognition of faces special vis-à-vis non-face objects is that the representation of faces retains aspects of combinations of their spatial frequency and orientation content extracted from earlier visual areas.”

Our study compared retaining (in phase-scrambled images) vs. removing (in wavelet scrambled images) the orientation information of the faces and means of transport images. Our results add a piece to the puzzle of what information is special about face processing: not only does a loss of orientation and phase information lead to impaired face recognition as compared to other objects as Yue et al. (2006) conclude. Additionally the presence of this information triggers an early face bias, even in the absence of semantic evidence (as our results indicate).

An event-related potential at 170-ms post-stimulus is widely thought to reflect early face processing mechanisms related to the structural encoding of faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bentin & Deouell, 2000; Eimer, 2000; Holmes, Vuilleumier, & Eimer, 2003) and is believed to be insensitive to changes in facial expression (Eimer, Holmes, & McGlone, 2003), although this last point has recently been brought into question (Schyns, Petro, & Smith, 2007).

We considered those saccades as “fast” that occurred earlier than the median reaction time of each subject. The average median reaction times were 176 ms in the phase group and 180 ms in the wavelet group with the earliest reactions around 120 ms.

Such early reaction times are incompatible with cortical detection mechanisms at 170 ms post-stimulus. Our RTs are similar to those found in recent ultrarapid categorization experiments with saccadic responses (Fletcher-Watson et al., 2008; Kirchner & Thorpe, 2006). Our results therefore highlight further a crucial question that these studies pose: Can a “high-level” cortical visual mechanism account for early reaction times at ~120 ms?

Note on the other hand that the early, low-level bias demonstrated here might not play a major role in relatively slower categorization processes: Earlier studies using manual responses did not find EEG correlates of natural scene categorization before 150 ms post-stimulus, and no selective reaction time shorter than 250 ms (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001b). Importantly, these categorization responses showed no preference for one target category compared to another (Rousselet, Macé, & Fabre-Thorpe, 2003; VanRullen & Thorpe, 2001a). Hence, the low-level face bias might be specific to experimental conditions that favor ultrafast responses, e.g., by using saccadic eye movements.

Fast responses toward phase scrambled images require a mechanism that works at least in part in a phase-invariant manner. Since the early 1970s, spatial frequency analysis has been applied to the description of receptive fields in primate and human visual pathways (Campbell & Robson, 1968; De Valois & De Valois, 1980; Shapley & Lennie, 1985). It had been assumed that the visual system up through V1 of macaques acts as a quasi-linear spatial filter over a wide range of contrasts (Braddick, Campbell, & Atkinson, 1978; De Valois & De Valois, 1980), thus performing in effect a local spatial frequency analysis. Complex cells in macaque striate cortex respond to drifting and counterphase flickering gratings of their preferred spatial frequency in a by and large phase-invariant manner (De Valois, Albrecht, & Thorell, 1982; Hubel & Wiesel, 1959). Up-ventral stream selectivities driven by complex cells should therefore maintain some activation in response to moderately phase-scrambled images, at least for the range of spatial frequency tunings found in striate complex cells. Striate simple and complex cells individually respond to a rather narrow range of spatial frequencies with a half-amplitude bandwidth between 0.5 and 2.0 octaves (De Valois et al., 1982). However their peak sensitivities cover a wide range of frequencies, from 0.5 c/deg up to 15 c/deg for retinal positions within 5 degrees visual angle from the fovea. This large bandwidth covered by striate complex cells...
would definitely suffice to maintain similar activity patterns in response to our phase scrambled face images.

V4, an area in the macaque ventral stream, receives direct input from V1 neurons and has been suspected to construct translation-invariant shape representations from its striate inputs in a feed-forward manner (Cadieu et al., 2007; Riesenhuber & Poggio, 1999). Area V4, as part of the ventral processing stream (Ungerleider & Haxby, 1994), is understood as an intermediate stage in the synthesis of complex visual objects. A recent proposal from our group (Kirchner & Thorpe, 2006) suggested that a rapid and parallel update of visual information between V4 and cortical eye fields such as FEF and LIP and/or superior colliculus could account for ultrarapid saccadic responses to visual targets. Stimuli of very high behavioral relevance such as faces or predators could possibly drive a fast saccadic bias through a short route of this kind, bypassing later stages of inferotemporal processing. In conclusion, we speculate that fast face biases to scrambled images, could depend on V4/cortical eye field activation by phase invariant cells in early visual areas.

**Perspectives**

Ultimate support must come from single cell recordings for the assumption of a cortical object detection pathway that relies principally on the amplitude spectrum of an object category. Furthermore, we exploited the tendency of humans to saccade to faces and used an indirect task (contrast detection). This allowed us to quantify the face bias as well as its behavior when introducing two different types of scrambling. However, we did not test for the same effects with other object categories than faces and means of transport. We conjecture that extensive training with another object category could produce a similar bias in non-semantic images as reported here for face images.

**Conclusions**

We performed a rapid categorization experiment, using eye tracking in a dual presentation paradigm. Subjects had a large individual and average bias toward the face images in a simple contrast judgment task. When scrambling both face and distracter images, by randomly shifting the phases of oriented spatial frequency components, the bias toward face images decreased but stayed surprisingly large with fully scrambled images, but only for fast saccadic reactions. Our results point to a rapid and automatic cortical mechanism based on activation of phase invariant cells in early visual areas and a rapid and parallel update of cortical eye fields through intermediate area V4 of the ventral pathway. Such a mechanism could direct attention to highly relevant stimuli, such as a guy next to his Jaguar, driven by the first wave of information reaching visual cortex.

**Acknowledgments**

Christian Honey has been supported by the German National Academic Foundation during his project in Toulouse. Rufin VanRullen is supported by the Fyssen Foundation, the Agence Nationale pour la Recherche (Grant 06-JC-0154-01), and the EURYI. The manuscript was improved by constructive comments from Guillaume Rousselet and an anonymous referee.

Commercial relationships: none.

Corresponding author: Rufin VanRullen.

Email: rufin.vanrullen@cerco.ups-tlse.fr.

Address: Centre de Recherche Cerveau et Cognition, Faculte de Medecine Rangueil, 31062 Toulouse Cedex, France.

**References**


Párraga, C. A., Troschianko, T., & Tolhurst, D. J. (2000). The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology, 10*, 35–38. [PubMed] [Article]


