Neural Correlates of Face Detection

Xiaokun Xu1 and Irving Biederman1, 2

1 Department of Psychology, 2 Neuroscience Program, University of Southern California, Los Angeles, CA 90089-2520, USA

Address correspondence to Xiaokun Xu, Department of Psychology, University of Southern California, 3620 McClintock Avenue, Seely G. Mudd Building, Room 501, Los Angeles, CA 90089-1061, USA. Email: xiaokunx@usc.edu

Although face detection likely played an essential adaptive role in our evolutionary past and in contemporary social interactions, there have been few rigorous studies investigating its neural correlates. MJH, a prosopagnosic with bilateral lesions to the ventral temporal-occipital cortices encompassing the posterior face areas (fusiform and occipital face areas), expresses no subjective difficulty in face detection, suggesting that these posterior face areas do not mediate face detection exclusively. Despite his normal contrast sensitivity and visual acuity in foveal vision, the present study nevertheless revealed significant face detection deficits in MJH. Compared with controls, MJH showed a lower tolerance to noise in the phase spectrum for faces (vs. cars), reflected in his higher detection threshold for faces. MJH’s lesions in bilateral occipito-temporal cortices thus appear to have produced a deficit not only in face individuation, but also in face detection.

Keywords: face detection, fusiform face area, phase-coherency-defined thresholds, prosopagnosia, saccadic choice

Introduction

Twenty thousand years ago, at low light, you are making your way through a dense forest inhabited by a hostile clan. Your likelihood of detecting a face peering out of the foliage could well determine your chance of survival. Given this evolutionary selective pressure for face detection, is there actually greater sensitivity for the detection of faces compared with other stimuli?

There is ample anecdotal evidence for a bias to interpret ambiguous stimuli as faces, as when we interpret clouds or rock formations as faces or the ambiguous Mooney faces compared with blobs. All these tend to form horizontally elongated, vertically aligned bands, mimicking (from the top) a 3 cycle/face barcode of the cheeks, dark (mouth), and light (chin). Sinha et al. (2006) also proposed a scheme that simply relies on the contrast regularity among the forehead, eye socket, and cheekbone. Similarly, a template consisting of a few dark and light rectangles has also been used for face detection in the initial screening in a popular computer vision model (Viola and Jones 2004). Since the early visual areas are tuned to the position and orientation of local contrast at multiple spatial frequencies (Hubel and Wiesel 1968), it is possible that face detection, that is, distinguishing a face from random natural noise, could be carried out solely from the information computed in this early stage.

On the other hand, extensive neuroimaging research has implicated regions in the later stage of ventral pathway that are specialized in the visual processing of faces. Specifically, the fusiform and occipital face areas (FFA and OFA, respectively) have been suggested as neural loci for the conscious perception of faces versus nonface objects: Higher activation in FFA has been found not only in viewing faces versus nonfaces (Puce et al. 1996; Kanwisher et al. 1997), but also in the conscious awareness of a face in binocular rivalry (Tong et al. 1998), in the figure/ground illusion (Hasson et al. 2001), as well as in face imagery without external visual input (Ishai et al. 2000; O’Craven and Kanwisher 2000). Grill-Spector et al. (2004) showed that the blood oxygen level-dependent signal in FFA was correlated not only with the successful identification of individual faces, but also with the simple detection of face, against nonface objects in brief, masked stimuli. Yue et al. (2006) reported a release from FFA adaptation when discriminating faces of different (vs. same) individuals but not when discriminating equally similar blobs, scaled by a measure of V1 similarity, designed to mimic the low-level features of faces. Andrews and Schlouppeck (2004) also reported higher activation in FFA associated with the awareness of ambiguous Mooney faces compared with blobs. All these findings strongly suggest that FFA is critically involved in the categorical and conscious perception of faces against nonface objects. However, it remains unclear whether these areas mediate face detection, as in the discrimination of faces against random noise.

Prosopagnosia, known as “face blindness,” is a disorder of face perception where individuals present severe deficits when individualizing faces. Depending on its origin, prosopagnosia could be classified as “congenital” (also termed “familial” or “developmental”) in the absence of apparent brain lesions (Duchaine et al. 2003). Numerous case studies of congenital prosopagnosia have reported normal face detection, with the majority reporting normal brain anatomy (de Gelder and Rouw 2000; Duchaine and Nakayama 2006; Garrido et al. 2008; Le Grand et al. 2006). Alternatively, prosopagnosia could be acquired, as a result of brain lesions, likely in the fusiform gyrus in the ventral temporal lobe (Damasio et al. 1990; Rossion et al. 2003). Most investigations of prosopagnosia have, understandably, focused on impairments in face identification or individuation, the core complaint of prosopagnosics. Little attention has been paid to their detection of faces.
To the best of our knowledge, there has not been a single report of impaired face detection in acquired prosopagnosia. In some of these cases, where there was a report of normal face detection, no lesion after recovery from closed head injury was observed (de Gelder and Rouw, 2000). In a few other cases (also with normal face detection), the face areas were only partially lesioned, such as in patient PS who suffered damage to both the left FFA and right OFA but with sparing of the right FFA (Rossion et al. 2003; Schiltz et al. 2006). If the posterior face areas, that is OFA and FFA, do mediate face detection, it may be the case that some surviving tissue in those areas is sufficient to achieve normal face detection. Would a more extensive lesion in those face areas impair face detection?

A potential shortcoming in previous investigations of face detection in prosopagnosia arises from their employment of designs in which detection of faces was compared with grid-scrambled faces, or 2-tone face contours embedded in a scramble of face features (Garrido et al. 2008). These types of stimuli do not readily permit principled parametric variation in the context of more general accounts of shape coding.

In the present study, in 2 experiments, we investigated face detection in MJH, an acquired prosopagnosic with extensive bilateral lesions to areas that would encompass the FFA and OFA in normal individuals, using more rigorous and theoretically motivated psychophysical tasks and metrics. Specifically, we manipulated the coherency of images by adding variable proportions of noise to the phase spectrum and measured the detection threshold—defined as the proportion of phase noise that allowed 75% of 2-AFC accuracy at distinguishing an instance of a target class (faces and cars) against a foil that was pure noise. In both experiments, the power spectra of the target and foil were identical in each trial. In Experiment 1, however, the power spectra of the faces and cars were unmatched, which allowed the possibility of an interaction between phase coherency and power spectra to affect performance. In Experiment 2, the power spectra were identical for the 2 stimulus classes.

Methods and Results

Subjects

MJH
In 1972 (40 years prior to the time of testing) at age 5, as the result of a fall from an 8-ft. high ledge, MJH suffered extensive bilateral lesions (greater in the right hemisphere) to his ventral occipito-temporal cortices, with extensive lesions in areas that would normally encompass FFA and OFA (Fig. 1, left panel). Anatomical inspection revealed no lesions in his superior temporal sulci. A contrast between face and nonface object presentation failed to reveal any significant activation in his ventral occipito-temporal lobe, even with a liberal threshold (Fig. 1, right panel) when he was tested in 2012 at the Dana and David Dornsife Cognitive Neuroscience Imaging Center at the University of Southern California, using the same block design FFA localizer employed by Xu et al. (2009). He does show normal activation to objects compared to their scrambles, in his spared lateral occipital cortex (LO), but not in the lesioned posterior fusiform gyrus (pFs).

Although there was a period of time immediately following his accident when he reported being completely blind, he regained close-to-normal vision, and currently exhibits normal contrast sensitivity as assessed with the Pelli-Robson Contrast Chart (Pelli et al. 1988), although he (inconsistently, over a number of years, in ophthalmic perimetry testing) sometimes presents some lower visual field loss in the periphery, particularly in the right visual field. On the Boston Naming Task (Kaplan et al. 1983), his performance is in the normal range (actually slightly above average with 47/50 correct) in identifying objects (Michel on and Biederman, 2003). He drives—in Los Angeles. To casual and nonrigorous examination, as well as subjective report, he is normal, or near normal, in his detection of faces. However, he shows pronounced impairment in individuating faces, on both standard tests such as the Benton Face Recognition Test 35/54 (Benton et al. 1983), the Cambridge Face Memory Test 42/58 (Duchaine et al. 2006), and a match-to-sample test in which an identical matching face is paired with a distracter face differing in identity (Yue et al. 2012). Mangini and Biederman (2004) reported that, on a test administered in 1999, he was at chance (controls were perfect) in selecting a celebrity (e.g. Bill Clinton) from a non-celebrity in pairs of faces, all of whom were highly familiar to him. In 2012, he was still at chance in individuating faces of celebrities in a similar choice test (again, controls were perfect). He can readily individuate a person on the basis of voice and shows normal, if not superior, memory for names and biographical details of the people he encounters. He is in the normal range in discriminating expression and sex (Mangini and Biederman, 2004) and reports that he has mental imagery (i.e. image memory) of faces that he has previously encountered (Michel on and Biederman, 2003). He is well aware of his deficit in individuating faces, reporting that he does not recognize his own face in a mirror nor those of close family members. However, he does not subjectively complain about face detection, instead reporting that all faces (within broad categories of age, sex, race, etc.) look the same. It is of some interest that, in the 40+ years since his accident at the age of 5, there has been insufficient plasticity to restore even a minimal ability to individuate faces.

Control Subjects

Control subjects (details specified separately for the 2 experiments) were comprised of students, faculty, and staff recruited from USC’s campus community. All controls reported normal or corrected-to-normal vision and no history of neurological disorders or injuries. The study was approved by USC’s Institutional Review Board.

Experiment 1: Detection Thresholds Defined by Phase Spectrum Coherency

Because shape identifiability is largely a function of the spatial phase (Oppenheim and Lim, 1981), by introducing external noise into the phase spectrum, the detectability of instances of stimulus classes (faces and cars in the present investigation) can be varied parametrically without affecting the global luminance and contrast of the images (Dakin et al. 2002; Sadj and Sinha 2004). Thirty-six individual Caucasian faces (half females) were created using the FaceGen Modeller 3.2 (Singular Inversions, Toronto, Canada, http://facegen.com), with moderate variation in their pose, size, and expression. Thirty-six car images were downloaded from the

Spectrum Coherency

Experiment 1: Detection Thresholds Defined by Phase Spectrum Coherency

Because shape identifiability is largely a function of the spatial phase (Oppenheim and Lim, 1981), by introducing external noise into the phase spectrum, the detectability of instances of stimulus classes (faces and cars in the present investigation) can be varied parametrically without affecting the global luminance and contrast of the images (Dakin et al. 2002; Sadj and Sinha 2004). Thirty-six individual Caucasian faces (half females) were created using the FaceGen Modeller 3.2 (Singular Inversions, Toronto, Canada, http://facegen.com), with moderate variation in their pose, size, and expression. Thirty-six car images were downloaded from the
internet, also with variation in their make, pose, and size (all 72 raw stimuli are shown in Supplementary Fig. 1). All images were converted to 8-bit, gray-scale images and then normalized in their global luminance histogram and the root mean square (RMS) contrast using the SHINE toolbox (Willenbockel et al. 2010). For each image, after a 2-dimensional Fourier transformation, we combined its phase spectrum with the counterpart of a white noise image with complimentary weights while maintaining its power spectrum. We adopted Dakin et al.’s method (2002) in the phase-blending process to avoid the over-representation of near-zero-degree phase components. The new image was produced by combining the original power and the synthesized phase spectrum through the reverse Fourier transformation, as shown in Figure 2A. A demo video showing a transition from 0% to 100% phase spectrum integrity of a face and a car can be downloaded at http://geon.usc.edu/~kun/PhaseModulationDemo.mov. The phase spectrum signal-to-noise ratio (psSNR) was defined as the ratio between the proportions of phase contributed by the original image (signal) and by the noise. Given the log-linear relationship between the psSNR and the detectability of the stimuli (Horner and Andrews 2009), the QUEST method (Watson and Pelli 1983) was used to probe the threshold of psSNR for 75% detection accuracy for faces and cars, respectively.

Participants performed a 2-AFC task where 2 images were presented side-by-side briefly either for 100 ms or for 200 ms, in separate blocks, and covered by grid-scrambled masks. Each image subtended a visual angle of approximately 3° and was centered at 2° eccentricity, left and right, from central fixation. Nineteen control subjects (10 females, mean age = 32.4 years, standard error of the mean, SEM = 4.1) and MJH

Figure 1. MRI brain imaging of MJH. (left) Ventral surface of MJH’s brain reconstructed from anatomical MRI scans with image provided by Hanna Damasio and Jessica L. Wisniewski from the Dana and David Dornsife Cognitive Neuroscience Imaging Center, USC, based on MR data obtained under P50NS019632. The blue colored patches (added by the authors) show the overlap between the FFA and OFA of controls and the lesion (cavities, shaded in white) in MJH. (right) The comparison of face-selective activation in controls and MJH based on a typical localizer functional scan. There is no preferential activation to faces, in contrast with objects, in MJH even with a liberal threshold.

Figure 2. Illustration of phase-coherence modulation for faces. (A) Example of how the images were generated by the introduction of random noise into the phase spectrum. (B) Examples of resultant images with variation in the proportion of the original phase. The identical procedure was used with car targets.
_pressed the left or right arrow key to indicate the target image (face or car, in separate blocks, each comprised of 72 trials, as illustrated in Fig. 2B) that were modulated by different proportions of noise (variable psSNR) in its phase spectrum, against a distractor composed of 100% phase noise. Subjects were instructed, prior to each block as to whether the target class would be faces or cars. The QUEST algorithm estimated the best psSNR for the target in the next trial by using Bayesian statistics based on previous trials. The terminating asymptotic psSNR was deemed as the detection threshold for each type of stimulus and exposure duration. Subjects heard a beep as error feedback and were permitted to take a break between blocks (Fig. 3).

Results
For controls, the detection threshold, defined in terms of phase SNR, was lower for faces than cars; and for the longer than the shorter exposure durations (Fig. 4A). A repeated-measures analysis of variance (ANOVA) revealed significant main effects of both exposure duration: \( F_{1,18} = 16.9, P < 0.001 \) and stimulus category: \( F_{1,18} = 9.7, P = 0.006 \), as well as their interaction: \( F_{1,18} = 5.3, P < 0.05 \). Pairwise comparisons further revealed lower thresholds for detecting faces than cars at both the 100-ms exposure: \( t_{(18)} = 4.3, P < 0.001 \) and the 200-ms exposure: \( t_{(12)} = 3.1, P = 0.006 \); as well as lower thresholds for the longer than the shorter exposure durations, for both faces: \( t_{(12)} = 2.3, P = 0.04 \) and cars: \( t_{(12)} = 2.7, P = 0.01 \). A further comparison of the exposure effects on the thresholds of faces and cars showed that the interaction was driven by a larger drop of car than face detection thresholds as a result of the longer presentation duration.

Compared with controls, MJH showed a markedly higher threshold in detecting faces (dashed line, Fig. 4A) at both the 100-ms exposure: \( t = 3.0, P < 0.01 \) and 200-ms exposure durations: \( t = 1.7, P = 0.05 \), and to a lesser extent in detecting cars at the 100-ms exposure duration, \( t = 1.7, P = 0.06 \), but not at the 200-ms exposure duration: \( t < 1, P > 0.1 \). The modified \( t \)-test proposed by Crawford and Howell (1998) was adopted to correct the bias introduced by the relatively small sample (<30) of controls. Although the absolute threshold difference between MJH and controls, defined by phase SNR, was larger for cars than faces (Fig. 4A, left panel), in units of the standard error of the face threshold in controls (shown by the miniscule error bars and calculated as a \( t \)-value), the deficit for MJH was greater for faces than for cars particularly at the 100-ms exposure duration.

**Experiment 2: Detection Threshold with Equal Power Spectra**

Although we equaledized both the global luminance and an RMS contrast among all stimuli in Experiment 1, there were differences in the power spectra between faces and cars that might have affected their detection thresholds. Therefore, in Experiment 2, the original power spectrum of each exemplar was replaced by the grand mean of the power spectra of all faces and cars in the stimulus set (Fig. 5A). Thus, in Experiment 2, the variation among the modulated images would be solely a function of their phase coherency. We analyzed the resultant detection threshold similarly as in Experiment 1. In addition, the results of controls in Experiments 1 and 2 were subjected to a mixed type, repeated-measures ANOVA, with exposure duration (100 vs. 200 ms) and category (face vs. car) as within-subject factors, and the image property (unmatched vs. matched power spectra between faces and cars) as the between-subject factor, given that different groups of controls were recruited in the 2 experiments. The result of this analysis would therefore help to clarify whether the categorical difference between faces and cars in their detection thresholds defined by phase coherence could be influenced indirectly by the power spectrum, even when the power spectra were always kept identical for both the target and the distractor by design. Another set of 19 control subjects (10

---

**Figure 3.** Illustration of experimental procedure, where the subject is to chose the image (left or right) containing the target category (in this case, a face) with the variable SNR in the phase spectrum against the 100% phase noise image. The 2 images share identical power spectra. The QUEST staircase procedure (Watson and Pelli 1983) was used to probe the threshold SNR—the ratio between original phase and random noise phase—to achieve 75% detection accuracy separately for each stimulus category and exposure duration.
females, mean age = 30.0 years, SEM = 3.5) participated in Experiment 2. The remaining procedures were identical to those of Experiment 1.

Results
Again, a lower detection threshold, in terms of psSNR, was evident for faces when compared with cars, and for the longer than the shorter presentation durations in control subjects (solid line in Fig. 4B). As in Experiment 1, a 2 stimulus category (faces and cars) × 2 exposure duration (100 and 200 ms) repeated-measures ANOVA in controls revealed significant main effects for stimulus category: $F_{1,18} = 74.6$, $P < 0.001$, and exposure duration: $F_{1,18} = 14.1$, $P = 0.001$, as well as their interaction: $F_{1,18} = 7.4$, $P < 0.05$. Pairwise comparisons revealed lower thresholds for detecting faces than cars at both the 100-ms exposure: $t_{(18)} = 8.4$, $P < 0.001$ and 200-ms exposure durations: $t_{(12)} = 7.1$, $P < 0.001$; as well as lower thresholds for longer than shorter exposures, for both faces: $t_{(12)} = 5.1$, $P = 0.006$ and cars: $t_{(12)} = 5.3$, $P = 0.004$. Similar to the results of Experiment 1, the interaction was driven by a relatively larger effect of exposure duration on the detection of cars compared with faces.

The equalization of image power spectrum did not significantly increase the threshold for the detection of faces, but did so for the cars. A mixed ANOVA was conducted with exposure duration and stimulus category as a within-subject factor and the spectral difference as a between-subject factor. There was a significant effect of both exposure duration: $F_{1,36} = 23.7$, $P < 0.001$ and stimulus category: $F_{1,36} = 40.9$, $P < 0.001$, and their interaction: $F_{1,36} = 12.6$, $P = 0.001$. More importantly, there was no significant interaction between category and power spectrum equality: $F_{1,36} = 9.9$, $P = 0.003$. A further independent $t$-test between the thresholds with and without equal power spectra was performed for face and car stimuli, collapsed over exposure duration. This analysis showed a significant effect of power spectrum equalization only for cars: $t_{(24)} = 3.9$, $P < 0.001$, not for faces: $t_{(24)} = 1.2$, $P > 0.2$, suggesting a minimal role of power spectra on thresholds for face detection, and a moderate role for car detection.

Compared with controls, MJH still showed a markedly higher threshold in detecting faces (dashed line in Fig. 4B) at both the 100-ms exposure: $t = 5.8$, $P < 0.01$ and 200-ms exposure durations: $t = 2.4$, $P < 0.02$, but close-to-normal thresholds at detecting cars at both exposure durations: $t < 1$, $P > 0.2$. In summary, even under power spectra equalization, MJH’s deficit for face detection remained, but not for car detection, as shown in the right panel of Figure 4B.

Age Effect and Individual Variance
To test for a potential age effect, we split the controls into 2 groups: 13 young (9 females, mean age = 23.3 years, SEM = 4.1) and 6 older (2 females, mean age = 49.7 years, SEM = 4.1) participants.
SEM = 2.3), approximately matched to MJH’s age, and performed the same analyses as reported above for all subjects. The results for both groups were highly similar to those when all controls were pooled: Controls’ detection thresholds were lower for faces than for cars, and lower for longer exposure durations than shorter ones. More importantly, MJH’s deficits in face detection remained significant compared with either age group: MJH versus 13 young controls, \( t > 2, \ P < 0.02 \) for both exposure, MJH versus 6 age-matched controls, \( t = 2.7, \ P < 0.02 \) for 100 ms and \( t = 1.2, \ P = 0.13 \) for 200 ms exposure. We further conducted a regression of the controls’ detection thresholds on age, for each combination of exposure, stimulus category, and power spectrum manipulations. All regressions showed a slope of the SNR as a function of age that was close to zero (\( \sim 0.01 \) increase in SNR units/year, \( P > 0.2 \)) except for the 200-ms presentation containing a car with unmatched (with respect to faces) power spectrum with a slope of 0.02 and \( P=0.05 \) (Figs 6 and 7, for Experiments 1 and 2, respectively). MJH’s face detection threshold was well above the regressive prediction of his age, whereas his threshold for car detection was among those for the controls.

The larger variance in car compared with face thresholds could be attributed to differences in both the variability of the stimuli and variability in the features relied on by different subjects, particularly with the cars. To prevent the subjects from using certain canonical templates for detection, we chose images of faces and cars of different identities/makes, and various poses and sizes. However, there still existed greater similarity among all faces than cars, due to the nature that all human faces are highly similar with each other. In debriefing subjects, all mentioned the usage of eyes as the most salient feature for face detection; whereas for car detection, different subjects relied on different features such as headlights, windows, and sharp angles in the contour, yielding considerable variability in their performance. In summary, the detection of faces seems more dependent on phase spectrum integrity, which was hardly affected by age. In contrast, the detection of cars appeared to rely not only on phase integrity, but also on the correspondence between phase and power spectrum. The explanation for this difference awaits further research.

**Discussion**

**Face Detection Thresholds**

By parametrically injecting noise in the phase domain, we degraded the detectability of images, somewhat mimicking the “walking in the woods” scenario where a target could be disguised by noise that resembles a phase-scrambled face or object. In control subjects, we found higher sensitivity to faces than cars, as evidenced by the lower detection threshold to faces. This is consistent with the previously reported saccade bias to faces (Crouzet et al. 2010), as well as the attention capture by faces (Vuilleumier 2000; Bindemann et al. 2005). What is the underlying mechanism supporting our higher sensitivity to faces? Torralba and Oliva (2003) suggested that the statistics of the power spectrum were sufficient to predict human detection of animals in natural images. However, Wichmann and Gegenfurtner (2010) reported that observers’ rapid detection of animals in the Corel image database to be essentially unchanged after power spectrum normalization among images of all categories. This result makes it highly unlikely that human observers make use of the global power spectrum. In the present study, could the lower threshold for faces versus cars be a function of a difference in the power spectra between the stimulus classes?

Strictly speaking, the power spectrum is not directly informative for choosing a target against noise in the present 2-AFC tasks, because these 2 always shared the identical power spectrum by design. However, the categorical difference in the power spectrum between faces and cars could still contribute to their detection threshold discrepancy by interacting with the phase spectrum modulation. We therefore
Figure 6. Regression of the detection threshold on age for controls in Experiment 1.

Regression of threshold on age (× MJH₉ controls), original power spectrum

Face 100ms

Car 100ms

slope < 0.01, P = 0.5

slope = 0.01, P = 0.3

Face 200ms

Car 200ms

slope < 0.01, P = 0.2

slope = 0.02, P = 0.05

Figure 7. Regression of the detection thresholds on age for controls in Experiment 2 (equal power spectrum for faces and cars).
further eliminated this possible confound by using the grand mean power spectrum of all stimuli in the phase-blending process, and then repeated the staircase procedure to measure the detection thresholds in Experiment 2. Controls showed again “lower” thresholds for faces than cars. This result suggests an intrinsic sensitivity to faces compared with nonface objects by human observers. Furthermore, this sensitivity is primarily driven by the local contours and features determined by the phase spectrum, whereas the integrity of the power spectrum plays only a minimal role. On the other hand, the detection of cars relied more on both phase spectrum coherency, and its interaction with the power spectrum structure even when the power spectrum by itself was not directly informative.

**Face Sensitivity Examined in Other Visual Tasks**

Hershler and Hochstein (2005) first reported “pop-out” in visual search for face targets among nonface objects such as vehicles and animals. Specifically, search inefficiency, defined as the slope of RT against set size, appeared to be much higher for nonface objects than face images. VanRullen (2006) re-examined this phenomenon by replacing the power spectrum of the target face and all distractors with that of various cars and found decreased search efficiency as a result of the mismatched power and phase spectrum of face images. Therefore, VanRullen concluded that face pop-out is not driven by high-level visual information such as contour localization determined by phase spectrum structure, but rather by low-level properties such as the power spectrum. This result is only in partial agreement with the present study: When the original power spectrum of the image was replaced by the grand mean power spectrum of 36 faces and 36 cars, control subjects’ detection thresholds measured as phase spectrum coherency were raised for car targets and, to a lesser degree and insignificantly, for face targets as well. The discrepancy between the VanRullen et al. and the present studies might be a function of the following factors: 1) The power spectrum was kept identical between the target and the distractors in the present study by design, therefore the discriminative information in the power spectrum could only be utilized via interaction with the modulation in the phase spectrum, whereas in the previous visual search experiments the power spectrum was directly informative; 2) The eccentricity of the face target. It has been shown that peripheral vision suffers from crowding and more distortion in the phase spectrum (Greenwood et al. 2009), compared with the relatively unlocalized power spectrum. Therefore, the alteration of the power spectrum could have been more disruptive when the face is presented in the periphery, as when they were randomly positioned in a large 8-by-8 search array in the visual search studies (Hershler and Hochstein 2005; VanRullen 2006); whereas, in the 2-AFC paradigm in the present experiment, both the target and the distractor were always presented close to foveal vision.

Honey et al. (2008) used a saccadic choice task to measure a bias toward faces. In every trial, 1 of the 2 images had a contrast value of 40% (the reference) and the other (the probe) a randomly determined level between 16% and 100%. They measured probe choice probability as a function of probe contrast, separately for faces and means of transport as the probe category, and fitted the 2 resulting curves with cumulative normal functions. The found that the equal probability of choosing the category was reached at a lower contrast level when the face was the probe, and at a higher contrast when the vehicle was the probe. This bias toward faces held even when the phase spectrum of the images was completely wiped out by random noise, such that subject could not recognize the object at all. The mean luminance and RMS contrast were matched between the faces and vehicles. However, when the 2 images had different contrast (in 80% of the trials over the whole experiment), the power spectrum was modified as a consequence as well; although the effect of contrast modulation did not necessarily eliminate the categorical information contributed by the power spectrum, it remained unclear whether the persistent bias toward faces could be completely attributed to the categorical difference in the power spectrum structure between faces and cars.

Crouzet et al. (2010) reported an ultra-fast (~100 ms) initial saccadic preference toward a face image when a face and a vehicle were simultaneously presented in the periphery, even when the subjects were instructed to fixate the vehicle. Crouzet and Thorpe (2011) further tested this saccadic bias toward faces when the power spectra were normalized among the faces and cars, just as we did in Experiment 2. They found a decreased but still significant tendency in saccading toward a face compared with a car, regardless of which one was instructed as the target. An even stronger test of the effect of power spectra has been made by swapping the power spectra between faces and cars. Subjects’ initial saccades were again biased to the hybrid of face phase and car power spectrum, indicating the dominance of phase spectrum over power spectrum information. Furthermore, the detrimental effect of power spectrum swapping was comparable for face and car targets, reflected in the comparable magnitude of increases in both reaction times and error rates for the 2 types of targets.

Importantly, the phase spectrum of the cars and faces remained intact in that study. In comparison, the present study was primarily concerned with the integrity of the phase spectrum, which turned out to dominate the detectability of both faces and cars: A higher signal-to-noise ratio in the phase spectrum composition led to higher detection accuracy. In addition, the contrast between Experiments 1 and 2 also revealed a minor role for power spectra, but only for cars, not for faces.

In an unpublished study, we replicated Crouzet et al.’s (2010) finding of a significant bias in saccades to faces against vehicles (with their power spectra unmodulated) in a group of control subjects. Unfortunately, MJH’s saccadic response accuracy to either face or car targets was at chance. Considering that the 2 images were presented simultaneously in the periphery (10° eccentricity from the central fixation) by design of the saccade paradigm, MJH’s chance-level performance was likely caused by his deficits in peripheral vision, confirmed by a recent perimetry test (performed by his ophthalmologist), although his foveal vision appears to be intact as assessed by visual acuity and contrast sensitivity tests that we administered.

In summary, the preferential saccade to a face is driven principally by its phase spectrum. However, when the coding of phase coherency is inefficient, either through experimental manipulation (Honey et al. 2008), or due to a cortical lesion (MJH), the power spectrum signature of faces can be utilized to accomplish face detection.
Neural Correlates of Face Detection

Could face detection be achieved in early visual areas without the involvement of posterior face-selective areas, such as FFA/OFA? We investigated this question in our study of MJH, who suffered lesions to those face areas but with sparing of early visual areas. When the power spectra for cars and faces were unmatched (Experiment 1), MJH’s threshold was higher than controls for both cars and faces, with his deficit more pronounced when detecting faces than cars. However, when the stimuli were equalized in their power spectra (Experiment 2), MJH’s deficits in face detection remained significant, whereas his car detection threshold was close to that of the normal controls. This dissociation in the effect of power spectrum equalization is in line with the neural tuning properties along the visual hierarchy: FFA/OFA has been shown to respond well to the modulation of phase, but largely unaffected by the variation in the power spectra of face images (Horner and Andrews 2009). In fact, those areas could be reliably localized to the modulation of phase, but largely unaffected by the integrity of the phase spectrum. MJH’s threshold was higher than controls over MJH. Because we only investigated one nonface category, cars, our finding of a greater reliance on phase integrity for the detection of faces means that we cannot exclude the possible importance of phase integrity for other nonface categories, such as human bodies, animals, tools, and scenes. Additional research is required to assess the relative importance of phase integrity for these other stimulus classes.

Conclusion

Prosopagnosics have been reported not to have difficulty in face detection, despite their pronounced deficits in face individuation. The present case study of prosopagnosic MJH nonetheless revealed a significant deficit in his detection of faces when assessed with rigorous psychophysical testing. The detection threshold for faces by controls was largely invariant with variations in the power spectrum, but heavily affected by the integrity of the phase spectrum. MJH’s heightened face detection threshold, together with his profound impairment in face identification, are likely consequences of his bilateral occipito-temporal lesions that encompass extrastriate, posterior face-selective areas. His largely spared early visual areas alone are insufficient to support normal face detection.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This study was supported by National Science Foundation (grants 0420794, 0531177, and 0617699) to I.B.

Notes

We thank Sébastien Crouzet for authorizing the use of his stimuli and protocols in the unpublished saccadic choice experiment, and Bosco Tjan for providing the eye tracker for that experiment and advising on the data collection and analysis of all the experiments. Conflict of Interest: None declared.

References


