Size determines whether specialized expert processes are engaged for recognition of faces

Nan Yang
Department of Medicine, University of British Columbia, Vancouver, Canada
Graduate Program in Neuroscience, University of British Columbia, Vancouver, Canada

Fakhri Shafai
Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, Canada

Ipek Oruc
Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, Canada

Many influential models of face recognition postulate specialized expert processes that are engaged when viewing upright, own-race faces, as opposed to a general-purpose recognition route used for nonface objects and inverted or other-race faces. In contrast, others have argued that empirical differences do not stem from qualitatively distinct processing. We offer a potential resolution to this ongoing controversy. We hypothesize that faces engage specialized processes at large sizes only. To test this, we measured recognition efficiencies for a wide range of sizes. Upright face recognition efficiency increased with size. This was not due to better visibility of basic image features at large sizes. We ensured this by calculating efficiency relative to a specialized ideal observer unique to each individual that incorporated size-related changes in visibility and by measuring inverted efficiencies across the same range of face sizes. Inverted face recognition efficiencies did not change with size. A qualitative face inversion effect, defined as the ratio of relative upright and inverted efficiencies, showed a complete lack of inversion effects for small sizes up to 6°. In contrast, significant face inversion effects were found for all larger sizes. Size effects may stem from predominance of larger faces in the overall exposure to faces, which occur at closer viewing distances typical of social interaction. Our results offer a potential explanation for the contradictory findings in the literature regarding the special status of faces.

Introduction

Recognizing a face is a challenging task, since each individual face differs only subtly from others. This can be appreciated by viewing in the inverted orientation; faces that are readily distinct when upright are less distinguishable upside down. This face inversion effect has been demonstrated for various aspects of face processing (Boutet & Faubert, 2006; Goffaux & Rossion, 2007; Leder & Bruce, 2000; Thompson, 1980; Yin, 1969). One particularly interesting aspect of the face inversion effect is perhaps not the fact that inverted faces are hard to tell apart, but rather that differentiating upright faces comes so easily. Several perceptual and cognitive mechanisms have been put forward in an effort to explain how the visual system achieves robust performance in face recognition.

One influential hypothesis suggests that human observers use expert neural processes that are specialized for the recognition of upright own-race faces (Farah, Wilson, Drain, & Tanaka, 1998; Moscovitch, Winocur, & Behrmann, 1997). On the other hand, nonface objects and inverted or other-race faces do not benefit from this special form of processing. Instead, these stimuli are believed to go through an alternate, general-purpose recognition route. Among the main lines of evidence for the idea of specialized face processing, often termed holistic or configural (Maurer, Grand, & Mondloch, 2002), are (a) the face inversion effect, showing disproportionately impaired recognition of faces when inverted compared to other nonface objects (Bartlett & Searcy, 1993; Rossion, 2008); (b) the part–whole advantage, showing enhanced recognition of face parts (e.g., eyes, nose) when viewed within the whole face rather than in isolation (Tanaka & Farah, 1993; Tanaka & Sengco, 1997); and (c) the composite face effect, where the upper half of a face is harder to

Citation: Yang, N., Shafai, F., & Oruc, I. (2014). Size determines whether specialized expert processes are engaged for recognition of faces. *Journal of Vision, 14*(8):17, 1–12, http://www.journalofvision.org/content/14/8/17, doi:10.1167/14.8.17.

doi: 10.1167/14.8.17 Received October 25, 2013; published July 22, 2014 ISSN 1534-7362 © 2014 ARVO

recognize when fused with a different-identity bottom half than when viewed separately (Young, Hellawell, & Hay, 1987).

Alternatively, other studies have found that the advantage upright faces enjoy may not necessarily stem from a qualitative distinction in the way these are processed (e.g., Martelli, Majaj, & Pelli, 2005; Sekuler, Gaspar, Gold, & Bennett, 2004). For example, two studies examined the spatial frequency bands used for face identification and found no differences between the upright and inverted conditions (Gaspar, Sekuler, & Bennett, 2008; Willenbockel et al., 2010). Also inconsistent with the holistic processing hypothesis, Gold, Mundy, and Tjan (2012) have shown that human observers' ability to identify a whole face is not superior to what is predicted by an optimal linear integration of information from individual features. Recent studies examining the relation between observers' competence in face recognition and the degree with which they processed faces holistically have found strong (Wang, Li, Fang, Tian, & Liu, 2012), partial (Richler, Cheung, & Gauthier, 2011), and no association (Konar, Bennett, & Sekuler, 2010) between the two.

With evidence accumulating on both sides of the argument, the controversy remains unresolved. One potential explanation for this apparent contradiction is that specialized configural processes may not necessarily be engaged in every setting featuring upright, own-race faces. Instead, it is possible that additional conditions are required. We hypothesize that the size of the face stimulus is a factor that critically determines whether expert face-specific processes are recruited. The known conditions, such as upright orientation and own race, necessary for engaging expert face processes are direct consequences of observers' exposure to faces, i.e., the face diet. Generally speaking, upright and own-race faces dominate the typical observer's face diet. Sizes that are predominantly featured in the face diet would also be expected to have a comparable influence on face processing. Face size, up to minor variations across individual faces, depends on the distance at which faces are encountered in daily life. Considering the social nature of the face stimulus, most of our exposure to faces, particularly regarding duration of exposure, presumably occurs in the context of social interaction. We might view a small face—e.g., of a passerby walking at a distance down the street—on the order of seconds to minutes. Larger faces are often viewed for much longer durations within conversational distances in the context of social communication and interaction with family, friends, and colleagues. Thus it is possible that face size, as a proxy for viewing distance, may play a significant role in face perception.

Evidence for size-dependent changes in processing of facial identity has been observed in peak spatial frequencies critical for recognition (Oruc & Barton, 2010). The typical finding for all stimulus categories tested to date is a positive linear relation between stimulus size and peak critical spatial frequency as specified relative to the object (e.g., cycles/image), including letters, words, novel shapes, and inverted faces (Chung, Legge, & Tjan, 2002; Chung & Tjan, 2009; Majaj, Pelli, Kurshan, & Palomares, 2002; Oruc & Barton, 2010; Oruc & Landy, 2009). One stimulus category—upright faces—is the exception. For upright faces, this typical linear increase occurs only for face sizes smaller than 4.7° of visual angle. Beyond this size boundary, peak spatial frequencies for face recognition flatten out and remain fixed around 8 cycles/face-width for larger sizes (Oruc & Barton, 2010). In addition, this distinctive face-specific effect of size on diagnostic spatial frequencies generalizes to faces of both Caucasian and Asian ethnicity, but only for own-race faces. Other-race faces, like the inverted faces, follow the typical linear trend for all sizes seen for nonface stimuli (unpublished data).

Taken together, these findings suggest that recognition of own-race upright faces involves two different processes based on size. At small sizes, one similar to that associated with the processing of nonface stimuli is used. On the other hand, large sizes engage a qualitatively distinct process seen only for upright own-race faces. Based on the spatial-frequency results (Oruc & Barton, 2010), this qualitative switch in processing occurs somewhere between 4° and 8° of visual angle. However, whether this switch in processing strategy corresponds to one between part-based general-purpose and holistic face-specific processes is unknown. In the present study, we examine this question. We reason that if specialized expert face processing applies only to faces larger than a particular size, then observers' face recognition performance with large faces should be superior compared to smaller faces.

One issue with comparing recognition performance across sizes is that size impacts the visibility of images at a basic level. For example, a small image may be harder to recognize simply because high-spatial-frequency components are too fine to see clearly. Thus, we examined efficiency of face recognition across face sizes that ranged between 1° and 10° of visual angle. Face recognition efficiency is defined relative to an ideal observer that takes into account the changes in visibility constraints due to size changes. To accomplish this, we measured each observer's contrast sensitivity function and incorporated it into an individualized ideal observer simulation we call a CSF-ideal observer (Chung et al., 2002; Oruc & Barton, 2010; Oruc & Landy, 2009). As a result, our CSF-ideal observer was limited by the same visibility constraints at each face size as the human observers. We compared upright face recognition efficiencies to those with inverted faces across the same size range. Based on our hypothesis, we predict an elevation of upright face recognition efficiency with increasing size. For inverted faces, we predict lower efficiencies across the entire size range without a discernible effect of size.

Methods

Subjects

Eight subjects (six women and two men, ages 21–36) with normal or corrected-to-normal vision participated in this study. All subjects completed 10 size conditions in the upright component of the experiment and 10 size conditions in the inverted face component of the experiment. In addition, all subjects underwent an hour-long session of contrast sensitivity testing. Each subject participated in four to six hour-long sessions completed on different days. The protocol was approved by the review boards of the University of British Columbia and Vancouver Hospital, and informed consent was obtained in accordance with the principles in the Declaration of Helsinki. All subjects except the authors were naïve to the purposes of this experiment.

Experimental setup

The experiment was performed on a computer with a Cambridge Research Systems (CRS) VSG 2/3 graphics card and Sony Trinitron 17-in. monitor (model GDM-200 PS). The display was gamma corrected using an OptiCAL photometer (Model OP200-E) and software provided by CRS. Mean luminance of the display was 40 cd/m^2. The experiment was programmed in Matlab (www.mathworks.com) using tools from CRS VSG Toolbox for Matlab and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Stimuli

Five female faces with neutral expressions were selected from the Karolinska Database of Emotional Faces (Lundqvist & Litton, 1998). Face images were converted to gray scale and resized using Adobe Photoshop CS 8.0 (www.adobe.com). The faces were seen through an oval aperture; they were aligned horizontally by centering the tip of the nose and vertically by aligning the pupil height across all faces No distinguishing marks were present on the faces, to avoid discrimination based on these markings. Ten sizes of each image were used. The sizes ranged from 1° to 10° per face-width in retinal size (corresponding to 77 to 768 pixels on the screen) at the constant viewing distance of 73 cm. Part 1 of the experiment used upright images of the five face stimuli, while Part 2 of the experiment used inverted versions of the same five face stimuli.

The root-mean-square (RMS) contrast for the face stimuli is defined as the standard deviation of luminance divided by mean luminance. The mean luminance was set to half maximum luminance, and the RMS contrast was set to 1 inside the oval aperture in order to maintain standard contrast across the images prior to experimental manipulation of the contrast in threshold measurements. In-house scripts in Matlab were used to generate the oval mask, align the faces horizontally and vertically, and adjust the luminance and contrast.

Noise

Upright faces: For the upright face component of the experiment, two noise conditions were used: white noise and no noise. Following Oruc and Landy (2009), we used low-pass-filtered Gaussian white noise with a high-spatial-frequency cutoff (15 cpd) as our nominal white noise to avoid clipping of pixels due to extreme values associated with unfiltered Gaussian masks. A large Gaussian white-noise mask (2048 × 2048 pixels, i.e., 16 times the area of a stimulus image) was low-pass filtered using a Butterworth filter with the squared gain function

$$G^2(f) = \frac{1}{1 + (f/f_c)^{2n}},$$

where f denotes spatial frequency and f_c denotes the cutoff frequency ($n = 5$). At each trial in the white-noise condition, the large mask was circularly shifted by random offsets vertically and horizontally. The top left quadrant of the large mask was then added to the stimulus to be displayed at that trial.

The noise level was set at a value sufficient to elevate thresholds while keeping task difficulty manageable for the subjects. For the smallest face size (1°), RMS contrast of the unfiltered white noise was 0.05, while for all other sizes (2°–10°), RMS contrast was 0.1. Note that the specific noise contrast used is arbitrary, as the human data are compared to the CSF-ideal observer that goes through the same task as the human at the same noise levels. We have specified an efficiency measure, termed high-noise efficiency (Pelli & Farell, 1999), which is independent of the specific noise level (Bennett, Sekuler, & Ozin, 1999; Legge, Kersten, & Burgess, 1987; Oruc, Landy, & Pelli, 2006; Pelli, 1981; Tjan, Braje, Legge, & Kersten, 1995). All subjects

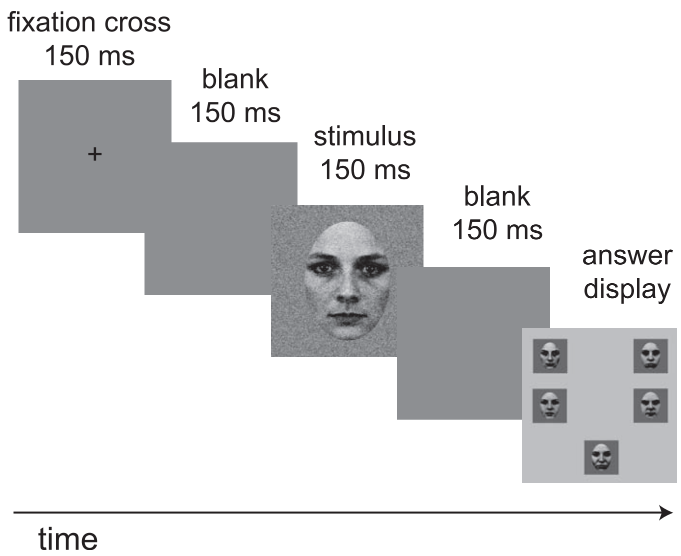

Figure 1. Sequence of displays in a typical trial. One of five possible face stimuli is shown for 150 ms either with no noise or with white noise. The observer chooses one of the choices on the answer display. This example is from Part 1, the upright component of the experiment.

completed the upright portion of the experiment using the noise levels as indicated.

Inverted faces: Similar to the upright face portion of the experiment, two noise conditions were used: white noise and no noise. The noise RMS contrast prior to filtering for the smallest face size (1°) of the inverted face was 0.025, while for all other sizes (2°–10°) it was 0.05 for all subjects with the exception of NY, who completed a more difficult version of the experiment with a noise RMS contrast of 0.025 for 1° and 0.1 for 2°–10°.

Procedure

We measured contrast thresholds at 82% accuracy for identification of face stimuli with and without added white noise in a five-alternative forced-choice paradigm. A trial of the experiment consisted of the following in sequence: a 150-ms fixation cross, a 150-ms blank, a 150-ms stimulus display, a 150-ms blank, and finally a choice screen displaying the five choices available, which remained visible until the observer entered a response (Figure 1). The responses were entered using keys on the computer keyboard that spatially corresponded to the position of the faces on the choice screen. An auditory feedback signal was provided with a single beep indicating a correct response and a double beep indicating an incorrect response.

The experimental trials were blocked according to face size, forming 10 blocks corresponding to the 10 face sizes (1°–10°). The subjects completed the blocks in a random order. Each block contained two randomly interleaved staircases, one staircase for the no-noise condition and the other for the white-noise condition, each lasting 40 trials. Staircases were implemented using the Quest procedure (Watson & Pelli, 1983) in Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for Matlab. A 40-trial training block was implemented for each face size, allowing the subjects to become accustomed to the size of the stimuli and the setting of experimental procedure. The experiment was completed with upright faces in Part 1 and inverted faces in Part 2.

Data analysis

Contrast thresholds

This experiment was conducted with two sets of stimuli: upright faces and inverted faces. All subjects completed two repetitions of the size blocks with each set of stimuli. Within each block, two contrast threshold estimates were obtained: one for the no-noise condition and a second for the white-noise condition. Threshold elevation was defined as the difference between squared threshold contrast (average of two estimates) in the white-noise condition and the no-noise condition. (See Supplementary Figure S1 for raw contrast thresholds for all subjects and conditions.)

CSF-ideal observer simulations

We compare human thresholds to those of an ideal observer to examine true human performance across stimulus sizes independent of intrinsic task difficulty. By definition, an ideal observer is a computer simulation that uses the least contrast energy possible to perform a given task (Geisler, 2011; Gold, Abbey, Tjan, & Kersten, 2009; Gold, Bennett, & Sekuler, 1999; Tjan et al., 1995). Thus the ideal observer provides a benchmark for human performance across conditions. In our particular case, the extraneous task difficulty we would like to remove from the raw data stems from low-level visibility changes due to changes in image size, which does not impact the performance of the standard ideal observer. Therefore, we used a CSF-ideal observer, a special form of the ideal observer simulation that takes the contrast sensitivity function of the human subject into consideration, generating an ideal observer with the same visual sensitivity and constraints as the corresponding human observer (Chung et al., 2002; Oruc & Landy, 2009). We measured individual contrast sensitivity functions (CSFs) for each observer. For each human CSF, we computed a unique equivalent input noise that was added to the face stimulus (see Appendix for details of the CSF experiment and the computation of the equivalent noise). This ensures that visibility of the image was similar for the human and the ideal observers across different sizes and that the performance of the CSF-ideal observer across sizes changed accordingly.

The CSF-ideal observer computation can be described as follows. The simulation completed 5,000 repetitions of an experimental block containing two 40-trial staircases, one for the no-noise condition and the other for the white-noise condition. At each trial, one of five possible face stimuli F_i ($i \in 1, \ldots, 5$) was selected randomly and presented to the ideal observer at the contrast c determined by the staircase. A noise mask N was then added to this stimulus, consisting of either the CSF-equivalent input noise (no-noise condition) or the sum of the white-noise mask and the equivalent input noise (white-noise condition, see Figure 2). The ideal observer computation was carried out in the Fourier domain to determine the face that was most likely to be the stimulus at each trial based on knowledge of the statistics of the total added noise, the five face templates, and the current trial contrast. The model accomplished this by selecting the face template that minimized the squared prewhitened difference between the noisy stimulus S and the template T_{i} as follows:

$$\min_{i=1..5} \sum \frac{(S - cT_i)^2}{N_{\mathrm{total}}}$$

where c is the trial contrast and N_{total} is the power spectrum of the total added noise, which in the white-noise condition was the sum of the external low-pass-filtered Gaussian white noise and the CSF-equivalent noise. In the no-noise condition, N_{total} was equal to the CSF-equivalent noise. For further details on the CSF-ideal observer computation, see Oruc and Landy (2009).

Efficiency

Efficiency was defined as the threshold energy elevation of the CSF-ideal observer divided by the human observer's threshold energy elevation in the same task, following Pelli and Farell (1999). Thus, efficiency E is calculated separately for each subject as

$$E = 100 \times \frac{c^2_{\mathrm{human},W} - c^2_{\mathrm{human},0}}{c^2_{\mathrm{ideal},W} - c^2_{\mathrm{ideal},0}}$$

where $c^2_{\mathrm{observer,noise}}$ is the squared contrast threshold for human and ideal observers in one of two external noise conditions (W = white noise, 0 = no noise). In addition to the external noise, the ideal observer also had the CSF-equivalent input noise associated with the corresponding human observer. The group efficiency was then computed as the geometric average of the individual efficiencies of all subjects. Note that this specific definition of efficiency, termed high-noise efficiency (Pelli & Farell, 1999), discounts the effects of internal noise and has been consistently shown to be invariant to external noise variance in many previous studies (e.g., Bennett et al., 1999; Legge et al., 1987; Pelli, 1981; Pelli & Farell, 1999; Tjan et al., 1995).

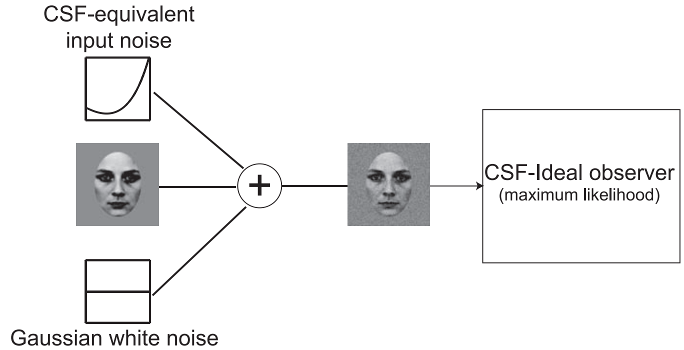

Figure 2. CSF-ideal observer is a maximum likelihood observer that takes the contrast sensitivity of the human subject into consideration through the addition of an equivalent input noise. A unique CSF-ideal observer is generated for each subject based on the subject's CSF results.

Statistics

Effects of orientation and size on efficiency were assessed using nonparametric Friedman ANOVA. One subject (JOL) showed a negative threshold elevation at the 3° inverted face condition due to contrast threshold estimates that were slightly higher in the no-noise condition than the white-noise condition. Thus there was a single data point out of 160 data points (8 subjects × 2 orientations × 10 sizes) with a negative efficiency value. This single data point was handled as follows: (a) For the purpose of statistical tests, we replaced this data point with the geometric average of the remaining seven subject's efficiencies at the 3° inverted face condition; and (b) for the plots of the data graphs (Figures 3 through 5), this single data point was left out of the geometric average for 3° inverted faces.

Results

Efficiency of recognizing upright and inverted faces as a function of face size is shown in Figure 3. There was a main effect of face orientation ($\chi^2 = 76.0500$, $df = 1$, $p \ll 0.001$), indicating significantly higher efficiencies in the upright condition (median = 1.5%) than in the inverted condition (median = 0.1%). Upright efficiencies showed a significant main effect of face size ($\chi^2 = 50.0727$, $df = 9$, $p \ll 0.001$), whereas for inverted efficiencies there was no main effect of size ($\chi^2 = 9.0818$, $df = 9$, $p = 0.4298$). For small upright faces

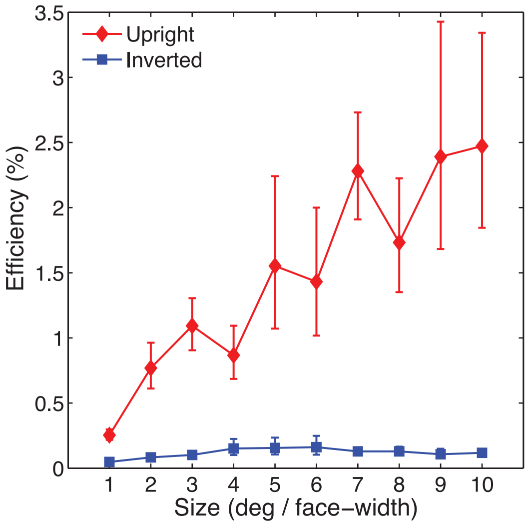

Figure 3. Experimental results. Efficiencies for upright and inverted faces are plotted as a function of face size. Group data were geometric averages across eight subjects; error bars show 68% bootstrap confidence intervals. Upright efficiencies were significantly higher than inverted. There was a significant main effect for face size in the upright condition but not in the inverted condition.

(2°–4°), efficiencies were around 1% or less and increased rapidly with further increases in face size to finally settle around more than double the initial efficiency. The curve for inverted faces, on the other

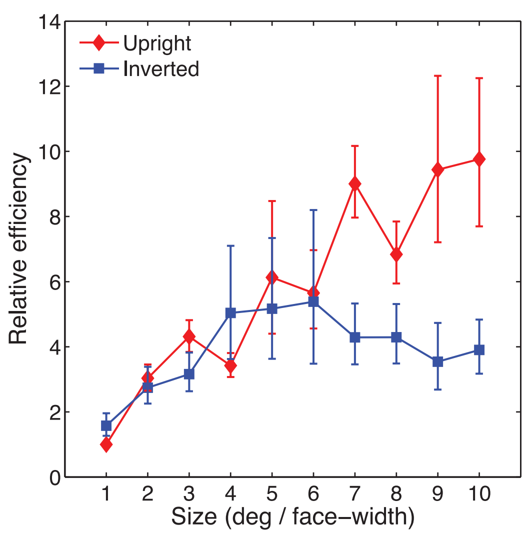

Figure 4. Relative efficiencies. To clarify qualitative effects of face size on the upright and inverted conditions, we normalized each efficiency curve to its corresponding minimum, thus removing the main effect between the two and retaining the relative changes in efficiency. The comparison between the two curves shows no discernible differences between the effect of face size on upright and inverted efficiencies up until 6° face size. After this size boundary, upright efficiencies continue to rise, whereas inverted efficiencies level off, marking a distinct effect of face size on the two orientations.

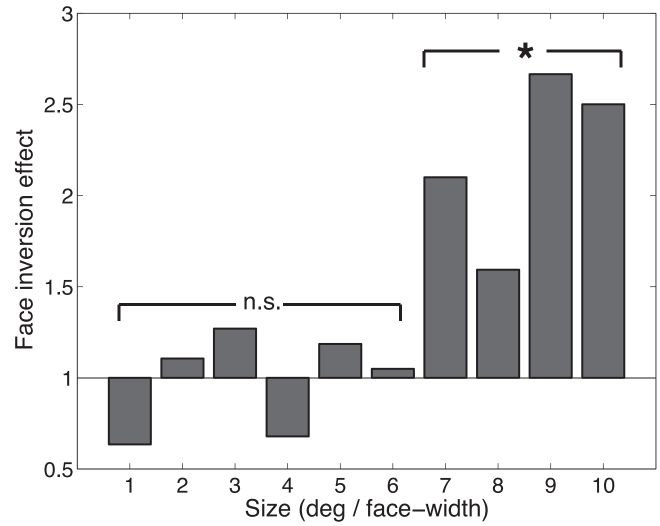

Figure 5. Qualitative face inversion effect. To quantify the distinct effect of face size on the upright and inverted efficiencies, we computed a qualitative face inversion effect (qFIE), defined as the ratio between the upright and inverted relative efficiencies. The null value for a lack of a qualitative inversion effect is a ratio of 1. Values above 1 signify the existence of qualitative face inversion effects. The qFIE did not differ significantly from 1 for sizes 1° to 6°, while for larger faces all qFIEs were significantly larger than 1 (all *ps* < 0.05) based on 95% bootstrap confidence intervals.

hand, stayed relatively flat, with some suggestion of an inverted-U-shaped pattern that peaked at 6° face size. However, this effect was not significant, as indicated by the lack of a main effect for size. One methodological shortcoming that may impact the interpretation of the differences between the upright and inverted face conditions is that these were not completed in a random order. In reviewing the potential order effects that might contribute, we argue that practice effects are unlikely, since inverted efficiencies were significantly lower than those at the upright condition despite the fact that the upright condition was completed first. Since upright and inverted sessions were completed on separate days, fatigue effects are also unlikely to contribute to the differences. Our main factor of interest, size, was completed in a random order, and thus order effects do not apply to the main effects of size. We cannot rule out the theoretical possibility that order of sessions might interact with the effects of size on efficiency. This is possible, but it is a fairly complicated hypothesis without a clear a priori reason as to why such a specific outcome would occur. Therefore we conclude that order effects do not provide a convincing alternative explanation for the differential effects of size across the two orientation conditions.

To examine the qualitative differences between the upright and inverted face data, we also calculate a *relative efficiency* by normalizing each series by its minimum efficiency, defined as

$$\left(\prod_{i=1..8}\frac{E_i(j)}{\min\limits_{j=1..10}E_i(j)}\right)^{1/8},$$

where $E_i(j)$ denotes efficiency of subject i at size j. Relative efficiencies (Figure 4) enable us to remove the overall difference in efficiencies between upright and inverted and to focus on the qualitative changes due to size. Relative-efficiency curves clearly show that the qualitative effect of size on efficiency is indistinguishable between upright and inverted conditions up until 6° face size. However, for faces larger than 6°, the two curves diverge. While upright efficiencies continue to rise with increasing size, inverted efficiencies stabilize or even decrease slightly.

To further examine the differences between upright and inverted relative efficiency curves, we compute a *qualitative face inversion effect* (qFIE), defined as the ratio between the two:

$$\text{qFIE} = \left(\prod_{i=1..8}\frac{E^{\text{upright}}{}_i(j)/\min\limits_{j=1..10}E^{\text{upright}}{}_i(j)}{E^{\text{inverted}}{}_i(j)/\min\limits_{j=1..10}E^{\text{inverted}}{}_i(j)}\right)^{1/8}.$$

A null value of 1 represents a lack of any qualitative face inversion effect. On the other hand, values larger than 1 signify qualitative effects of inversion. To determine the range of face sizes where the qFIE emerges, we performed individual t tests based on 95% bootstrap confidence intervals. For face sizes between 1° and 6°, qFIEs did not differ significantly from the null value of 1. In contrast, qFIEs were significantly larger than 1 (all ps < 0.05) for all larger faces (Figure 5).

Discussion

Many influential models of human face recognition postulate some form of specialized configural processing distinct from general-purpose part-based schemes. These expert processes are engaged by upright own-race faces, but not by inverted or other-race faces or nonface stimuli. On the other hand, there is also compelling evidence suggesting that the difference between processing of faces and other stimulus categories is a quantitative, not a qualitative one. The controversy remains.

Thus far, the size at which a face is viewed has not received much attention in the literature. Size, by and large, is considered to play little, if any, role in the processes that enable recognition of faces. This is a reasonable, albeit largely untested, view, since the size of a face contains no information regarding its identity. Our recent work (Oruc & Barton, 2010), however, on the role of spatial frequencies in facial identity recognition shows a qualitative switch from one distinct mode of processing to another between 4° and 8° face size. Specifically, this work shows that large faces were processed using relatively lower frequencies than what would be expected from an extension of the processing of smaller sizes. These results bring into light the possibility that size may be a determining factor in whether or not specialized holistic processes will be engaged. We hypothesized that expert holistic processing may be reserved for faces larger than a set size.

The purpose of the present study was to examine any fundamental changes in the face recognition ability of human observers as a function of size. We reasoned that if expert processes kick in once a set size boundary has been crossed, this would be reflected in superior face recognition abilities at those larger face sizes. In order to isolate face-specific changes in performance as a function of size from those that are due to changes in physical image visibility, we specified observers' efficiency of identification. Efficiency is a measure of performance relative to that of an ideal observer. Following previous work, we employed a specialized ideal observer, a CSF-ideal observer that incorporated the contrast sensitivity function of each human observer in the form of equivalent added noise (Oruc & Barton, 2010; Oruc & Landy, 2009). A unique CSF-ideal observer was created for each human observer who participated in the experiment. Thus, the individualized CSF-ideal observer operated under the same visibility constraints as the corresponding human observer and, consequently, controlled for changes in performance that are purely due to changes in visibility. A second way we isolated face-specific performance from lower-level factors is comparing upright face recognition efficiencies to those for inverted faces. Since upright and inverted face stimuli were identical except for orientation, any differences in efficiencies as a function of size must stem from the way in which the human brain processes these two stimuli.

Consistent with our hypothesis, upright face recognition efficiencies increased with size (Figure 3). This increase is fairly gradual and at first sight may appear to lack the steep localized slope that would be expected from a qualitative switch in processing at a set size. However, the true significance of this gradual increase cannot be appreciated without considering additional effects of increasing size on efficiency. Human observers, but not ideal observers, experience decreases in contrast sensitivity going from center to periphery, as well as an inability to integrate information across distances in a large image. Therefore, everything else being equal, efficiencies are often found to decrease with size (Kersten, 1984; Pelli, Burns, Farell, & Moore-Page, 2006; Tjan et al., 1995). In contrast, we observed an increase in efficiencies rather than the expected decrease. This increase in upright efficiency presumably represents the net effect that overrides this undercurrent of decline. In order to elucidate the true effect of size on face recognition abilities, we consider the upright face efficiencies in comparison to the efficiencies in the inverted condition. As expected, inverted efficiencies were substantially lower than upright. This is consistent with the well-established face-inversion effect, reflecting impaired performance with inverted faces compared to upright. To compare the qualitative changes in efficiencies of the two orientations, we defined a relative efficiency that was normalized to the minimum efficiency in each series (Figure 4). This enabled us to remove the overall difference in efficiencies between upright and inverted and to focus on the qualitative changes due to size.

Visual inspection of the relative-efficiency curves makes it clear that the effect of size on upright and inverted conditions is highly similar up until 6° face size. Beyond this point, the two curves diverge: The upright condition continues on an increasing trajectory, while the inverted efficiencies flatten out and even show trends of tapering off. For a closer inspection of these differences, we derived a qualitative face inversion effect (qFIE), defined as the ratio between upright and inverted efficiencies that were normalized to the minimum efficiency in each series (Figure 5). The qFIE as a function of size appears to be a step function abruptly switching at the 6° size boundary from a complete lack of qualitative differences to significant qualitative inversion effects.

Taken together, these results point to the conclusion that while upright faces are recognized more efficiently than inverted faces, the difference remains a quantitative one up until a size boundary of around 6°. Qualitative differences emerge only beyond this size boundary. Thus, upright faces take on the status of special stimulus category only at large sizes beyond 6° of visual angle. Our results offer a potential resolution to the apparently contradictory nature of past findings. We reviewed the face sizes used in 23 influential studies of face perception that argue for (16 out of 23) and against (7 out of 23) qualitative differences and specialized configural processing for faces. The results of this review, which includes 30 distinct face sizes used in these studies (Figure 6), show that studies that found evidence for specialized/configural processes used larger faces (red, 19 distinct sizes, median = 6°) than studies that did not find any qualitative differences (blue, 11 distinct sizes, median = 3.75°) that and this difference was significant ($p < 0.001$, two-sided) based on a nonparametric Wilcoxon rank-sum test. Thus, the controversy in the field regarding the processing of faces appears to stem in part from the choice of stimulus size.

What is the significance of the face size boundary of 6°? In daily interactions, faces at these large sizes are viewed when standing closer than 2 m to another person. Such interpersonal distances are typical in the context of conversations and close-phase social interactions. Since faces are often encountered in social situations, it is possible that the majority of our exposure to faces occurs at these close distances. It is generally accepted that expertise in faces is deeply shaped and determined by the face diet of individuals. For example, face expertise only applies for upright own-race faces, due to disproportionate exposure to such stimuli compared to inverted and other-race faces. A similar mechanism may be at play in bringing about the effects of size, if exposure to larger faces seen up close exceeds that for smaller faces seen from afar. Indeed, recent evidence suggests that this is the case for the face diet of infants. Sugden, Mohamed-Ali, and Moulson (2014) examined daily exposure to faces of 1- and 3-month old infants through head-mounted cameras. Although face size is not among the variables reported in this article (Sugden et al., 2014), it was nevertheless coded. The authors found that faces smaller than 5° comprised only 12% of total duration of face exposure—88% percent of duration occurred with larger faces (personal communication). Similarly, Smith and colleagues have recorded face exposure of 22 1- to 12-month old infants and found that faces were viewed predominantly at close distances between 60 cm and 120 cm (personal communication with L. Smith, March 15, 2014). This means within the first year of life, faces are viewed at sizes predominantly between at least 6° and 12°, based on a median female face width of 13.3 cm (bitragion breadth, Poston, 2000)—median male face is wider (14.5 cm), corresponding to a size range between approximately 7° and 14°. At this time, we know of no studies examining face exposure of adults; thus it remains to be seen whether the dominance of large faces in the face diet continues in adulthood.

Our present results point to superior performance in the identification of faces larger than 6°, although the qualitative changes in the processing of facial stimuli that give rise to this effect remain unclear. Holistic or configural processing, as opposed to part-based, has often been equated with specialized expertise for faces. Future studies should examine whether hallmarks of configural processing such as the part–whole advantage (Tanaka & Farah, 1993; Tanaka & Sengco, 1997) vary with face size. Peterson and Eckstein (2012) have found that human observers fixate a location just below the eyes that is statistically optimal for identification based on viewing at a normal conversational distance. Indeed, when forced to fixate elsewhere, observers' performance deteriorated. Two limitations of our study—lack of measurements of eye position and of peripheral CSFs for each individual subject—prevent us from

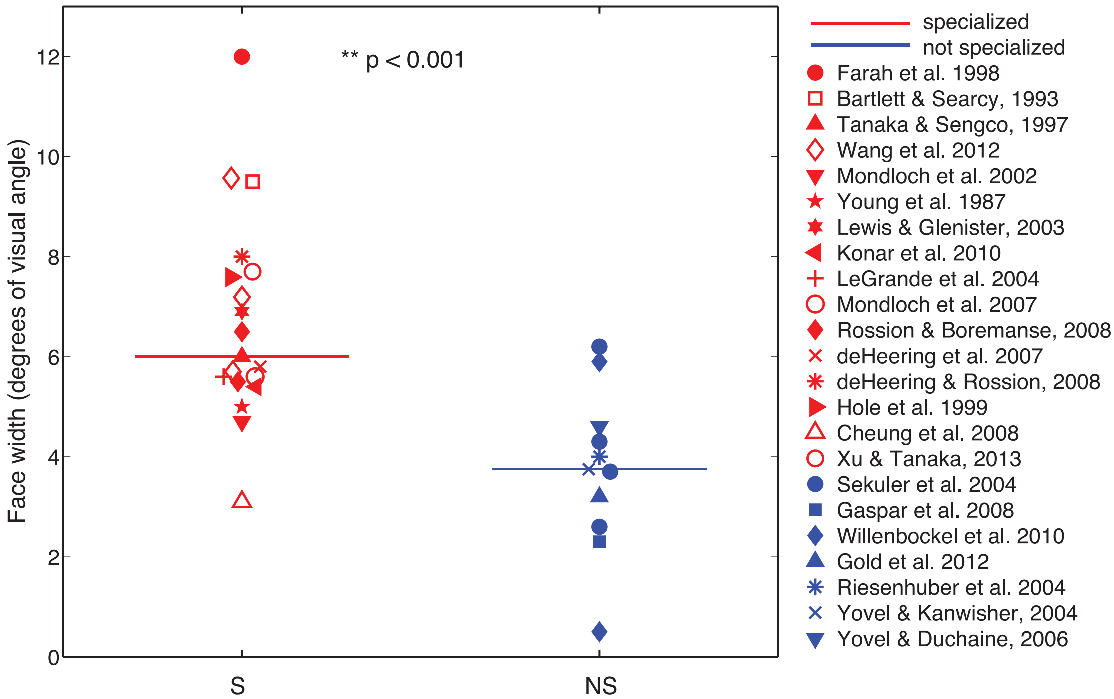

Figure 6. Sizes of face stimuli used in the literature. Face width in degrees of visual angles used in 23 influential studies of face perception is grouped based on whether evidence for qualitatively distinct processing specialized (*S*) for faces was found (red, 16 studies, 19 distinct sizes used) or whether such effects for specialized mechanisms were not (*NS*) evident (blue, seven studies, 11 distinct sizes used). Median face widths shown by the horizontal rules in each group (*S*: 6°; *NS*: 3.75°) differed significantly ($p <$ 0.001, two-sided Wilcoxon rank-sum test). Markers are jittered slightly to minimize overlaps.

examining whether changes in fixation strategies across sizes as well as orientation of faces are behind our present findings. A future study with measurements of fixation locations combined with ideal-observer modeling that takes into account changes in contrast sensitivity with eccentricity (Najemnik & Geisler, 2005; Peterson & Eckstein, 2012) can clarify this issue.

In conclusion, regardless of the origins of the effects of size that we have demonstrated, these findings offer a potential resolution to the ongoing controversy on the processing of faces. Our present results suggest that faces may be processed through generic recognition pathways or engage specialized processes depending on the size at which they are viewed. Future work is needed to further test this hypothesis. Our results apply to identification of static faces with no other external features or cues available to the observer. Thus, they do not necessarily extend to person identification in the real-life setting where observers may compensate by using other sources of static or dynamic information available to them, such as the shape of the body or characteristic movements of the body (e.g., gait). It is clear, however, that size must be taken into consideration as a significant factor in future studies of face perception.

Keywords: face recognition, size, scale invariance, efficiency, contrast threshold, ideal observer, configural processing, holistic, piecemeal

Acknowledgments

This work was supported by a Natural Sciences and Engineering Research Council of Canada, Discovery Grant RGPIN 402654-11 (IO). NY was supported by the University of British Columbia Summer Student Research Program.

Commercial relationships: none.
Corresponding author: Ipek Oruc.
Email: ipor@mail.ubc.ca.
Address: Department of Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, Canada.

References

Bartlett, J., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*, 281–316.

Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 16*(3), 654–668.

Boutet, I., & Faubert, J. (2006). Recognition of faces and complex objects in younger and older adults. *Memory & Cognition, 34*(4), 854–864.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Chung, S. T., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research, 42*(18), 2137–2152.

Chung, S. T., & Tjan, B. S. (2009). Spatial-frequency and contrast properties of reading in central and peripheral vision. *Journal of Vision, 9*(9):16, 1–19, http://www.journalofvision.org/content/9/9/16, doi:10.1167/9.9.16. [PubMed] [Article]

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482–498.

Gaspar, C., Sekuler, A. B., & Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vision Research, 48*(28), 2817–2826.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*(7), 771–781.

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 995–1002.

Gold, J., Abbey, C., Tjan, B. S., & Kersten, D. (2009). Ideal observers and efficiency: Commemorating 50 years of Tanner and Birdsall: Introduction. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 26*(11), IO1–2.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*(21), 3537–3560.

Gold, J., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science, 23*(4), 427–434.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24*(12), 1977–1990.

Konar, Y., Bennett, P. J., & Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychological Science, 21*(1), 38–43.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology: Section A, 53*(2), 513–536.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, 4*(2), 391–404.

Lundqvist, D., & Litton, J. E. (1998). The Averaged Karolinksa Directed Emotional Faces - AKDEF [CD-ROM]. Stockholm, Sweden: Department of Clinical Neuroscience, Karolinska Institutet.

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research, 42*(9), 1165–1184.

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1):6, 58–70, http://www.journalofvision.org/content/5/1/6, doi:10.1167/5.1.6. [PubMed] [Article]

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555–604.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391.

Oruc, I., & Barton, J. J. (2010). Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces. *Journal of Vision, 10*(12):20, 1–12, http://www.journalofvision.org/content/10/12/20, doi:10.1167/10.12.20.

Oruc, I., & Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *Journal of Vision, 9*(9):4, 1–19, http://www.journalofvision.org/content/9/9/4, doi:10.1167/9.9.4. [PubMed] [Article]

Oruc, I., Landy, M. S., & Pelli, D. G. (2006). Noise masking reveals channels for second-order letters. *Vision Research, 46*(8–9), 1493–1506.

Pelli, D. G. (1981). *Effects of visual noise*. Cambridge, United Kingdom: Cambridge University.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *16*(3), 647–653.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA, 109*(48), E3314–3323.

Poston, A. (2000). *Human Engineering Design Data Digest*. Retrieved from http://hfetag.com/hfs_docs.html

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science, 22*(4), 464–471.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*(2), 274–289.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*(5), 391–396.

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology, 56*(2), 249–261.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Section A, 46*(2), 225–245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*(5), 583–592.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9*(4), 483–484.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research, 35*(21), 3053–3069.

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science, 23*(2), 169–177.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33*(2), 113–120.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., . . . Gosselin, F. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 122–135.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759.

Appendix

The contrast sensitivity experiment

We measured the contrast sensitivity function (CSF) of each observer who participated in the face recognition experiment. This was done to characterize the changes in visibility introduced due to changes in the size of the face stimuli. We estimated an equivalent input noise spectral density based on the contrast thresholds measured as a function of spatial frequency. The individual equivalent input noise profiles allowed us to compute a unique CSF-ideal observer for each subject.

Methods

Subjects

The same group of subjects who completed the face recognition experiment participated in this experiment.

Stimuli

Stimuli were Gabor patches with six spatial frequencies: 0.5, 1, 2, 4, 8, and 16 cpd. We used vertical sine-phase Gabors with 1.2 octaves full bandwidth at half height.

Procedure

We measured contrast thresholds for detection in a two-interval forced-choice paradigm (2-IFC). Each trial started with a 150-ms fixation cross and a 150-ms blank. This was followed by two 150-ms intervals marked by temporally synchronous beeps and separated by a 150-ms blank. The Gabor patch was displayed at the center in one of the two intervals chosen randomly at each trial. The other interval contained a blank. After the second interval a blank was displayed until the subject responded by pressing the “1” or the “2” key on the computer keyboard to indicate the temporal interval that contained the stimulus. The contrast of the Gabor was controlled by two randomly interleaved staircases. Each session was blocked by spatial frequency, lasting 80 trials each (40 trials per staircase). The order of the six spatial-frequency blocks was randomized. Each subject completed two sessions, yielding four threshold estimates for spatial frequency.

Data analysis

Contrast threshold for each spatial frequency was based on the average of the four independent estimates obtained over two sessions. Since the Gabor patches varied in size with varying spatial frequency, we computed threshold signal energy $E(f)$ as squared contrast integrated over the stimulus area. Equivalent noise power at each spatial frequency was computed as

$$N_{eq}(f) = \frac{E(f)}{(d')^2},$$

where d' is the value of the d' corresponding to the 82% threshold criterion for the 2-IFC contrast detection task. We fitted a smooth parametric curve to equivalent-noise power estimates as a function of spatial frequency, which allowed us to interpolate between tested spatial frequencies and extrapolate beyond the lowest (0.5 cpd) and highest (16 cpd) spatial frequencies. For simplicity, we assumed sensitivities to be independent of orientation and eccentricity.