Critical frequencies in the perception of letters, faces, and novel shapes: Evidence for limited scale invariance for faces

İpek Oruç

Human Vision and Eye Movement Laboratory, Departments
of Medicine (Neurology) and Ophthalmology and Visual
Sciences, University of British Columbia,
Vancouver, BC, Canada

Jason J. S. Barton

Human Vision and Eye Movement Laboratory, Departments
of Medicine (Neurology) and Ophthalmology and Visual
Sciences, University of British Columbia,
Vancouver, BC, Canada

Despite the common intuition that object recognition processes should be relatively scale invariant, a number of studies show that this is not the case. Using a critical-band masking paradigm, we examined the pattern of scale dependence of diagnostic spatial frequencies across a range of stimuli that varied in participants' prior experience and the 'ecological significance' of the stimuli, by which we mean the degree of universality and recency of the development of the stimulus in human culture, letters being an example of a culturally arbitrary stimulus and faces a universal one. We found scale dependence for letters, mirror-image letters, and novel shapes, consistent with prior results, as well as for inverted faces. However, upright faces showed a relatively scale-invariant pattern especially for face sizes that corresponded to those encountered in typical social interactions. This suggests an important difference between the processing of faces and other objects that may reflect their unique status as stimuli.

Keywords: face recognition, scale invariance, object recognition, critical frequencies, critical-band masking, discrimination, contrast thresholds, holistic, part based

Citation: Oruç, İ., & Barton, J. J. S. (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.

Introduction

Object recognition is one of the most complex tasks the visual system faces. Images of objects undergo severe transformations due to variations in location, size, orientation, and illumination. This presents a challenge to any recognition algorithm, whether biological or computational, as recognition of the object should be invariant to image variations that reflect viewing conditions, while staying sensitive to those image properties that reflect the difference between different objects. Human object recognition is exquisitely robust in this respect. Consider the font and size changes one faces in reading. For instance, the two characters, 𝒜 and **A** are readily identifiable as the same letter even though the corresponding physical stimuli differ substantially.

Invariance to size has been demonstrated for various aspects of human object recognition. For example, training with different-size exemplars provided similar benefits in an object naming task as training with same-size stimuli (Furmanski & Engel, 2000). Another study of object naming found that the magnitude of a priming effect did not depend on whether the sizes of prime and test stimuli matched (Biederman & Cooper, 1992). Efficiency of letter identification and reading rate are only weakly affected by changes in letter size (Legge, Pelli, Rubin, & Schleske, 1985; Parish & Sperling, 1991; Pelli, Burns, Farell, & Moore-Page, 2006).

The fact that the visual performance of human observers is found to be scale invariant can be interpreted as indicating that the underlying recognition processes must also be scale invariant. However, recent evidence suggests otherwise. Majaj, Pelli, Kurshan, and Palomares (2002) have shown that the critical band of spatial frequencies for recognizing letters changes with letter size. Large letters are recognized with their details (higher frequency components) whereas small letters are recognized with their large strokes (lower frequency components), a finding that has been replicated by others (Chung, Legge, & Tjan, 2002; Oruc & Landy, 2009).

If object-recognition processes are inherently scale dependent, why do we not notice this in our everyday visual experience? The hybrid images of Oliva, Torralba,

doi: 10.1167/10.12.20 Received April 12, 2010; published October 18, 2010 ISSN 1534-7362 © ARVO

and Schyns (2006) provide an example where the scale-dependent nature of object perception is evident. Hybrid images are composites of multiple visual objects, each occupying a separate frequency band. Scale-invariant processing would mean that the percept of hybrid images would be the same regardless of image size. In reality, at different sizes different components of the hybrid image dominate the overall percept. For example, in Figure 1 most viewers report seeing Botticelli's Venus in the large image at the top, and the iconic "Love" image by Robert Indiana in the small image at the bottom. Actually, these two images are identical, simply printed at different sizes, which can be verified by standing 3–5 m away from the page or screen and observing that *Love* replaces *Venus* in the larger top image at this far viewing distance. Designing hybrid images that work requires knowledge of preferred, or critical, frequency bands for the component visual objects at various sizes and provides phenomenological evidence of scale dependence in our object recognition system. Such demonstrations are valuable for illustration purposes but do not provide proof of scale dependence. The most convincing evidence instead comes from systematic psychophysical experimentation (Chung et al., 2002; Chung & Tjan, 2009; Majaj et al., 2002; Oruc & Landy, 2009).

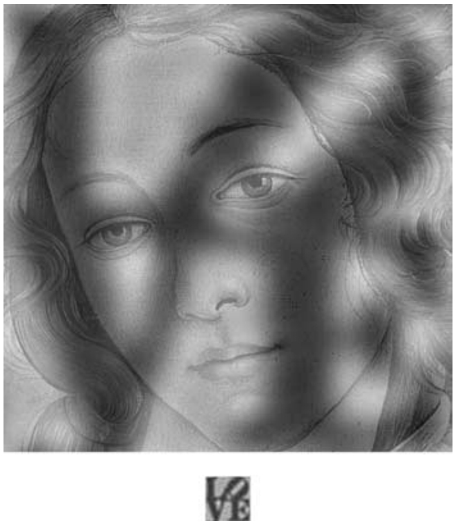


Figure 1. A hybrid image. Most people perceive the Venus of Botticelli at the top and the iconic Love image by Robert Indiana at the bottom, despite the fact that the bottom image is the same as the top image displayed at a smaller size. The reader is invited to view the top image from a distance to confirm that the dominant percept changes at this distance from *Venus* to *Love,* thus verifying that the different percepts in the top and bottom images are due to difference in size, not due to a printing artifact. This occurs because the scale at which an image is observed has an impact on what is perceived in that image. The hybrid image is designed to highlight this inherent scale dependence in human object recognition.

It was initially assumed that scale dependence reflected constraints on visual contrast sensitivity in human observers (Chung et al., 2002; Oliva et al., 2006; Oruc, 2003; Oruc, Landy, & Pelli, 2006). This account is based on the shape of the contrast sensitivity function (the CSF-based account), i.e., the fact that human sensitivity for very low and high spatial frequencies is limited compared to middle frequencies. Changes in object size can thus render some frequency components of an object hard to detect. For example, higher frequency components of a letter may be easily discerned when the letter is large, but when the letter is small these components will be located in even higher spatial frequencies as far as the retinal image is concerned, frequencies to which humans are far less sensitive. The pattern of scale dependence in human observers is in qualitative agreement with the predictions of the CSF-based account, in so far as the changes in preferred frequencies are in the expected direction (Chung et al., 2002; Oruc, 2003; Oruc et al., 2006). However, this account has recently been challenged. For example, one can make contrast sensitivity relatively equivalent across all spatial frequencies, i.e., obtain a considerably flatter CSF, by the addition of external white noise. Insensitivity to a particular spatial frequency is often modeled through the presence of higher internal noise in the neural processing of that stimulus (Ahumada & Watson, 1985). In other words, for higher and lower spatial frequencies, internal noise exceeds that for the middle frequencies. Consequently, the addition of high-power external white noise to the stimuli swamps the internal noise and the relatively minor differences in the internal noise become negligible. As a result, thresholds for all frequencies are raised considerably, and the characteristic shape of the CSF is rendered relatively flat. If the CSF-based account is correct, addition of external white noise should eliminate scale dependence, but it does not (Oruc & Landy, 2009), a finding that demonstrates that scale dependence has a deeper origin than low-level constraints on contrast sensitivity.

Many of these observations on scale dependence derive from experiments using letters as stimuli. One important question is whether such results generalize to all other objects, or if there are some fundamental differences between certain types of objects that may affect the results. Letters form an interesting class of stimuli. Although letters and written text may have been designed and in time tailored to broadly suit basic human visual capabilities, they remain an artificial class with which humans develop an arbitrary expertise that is culturally determined. Whether one develops an expertise for English or Korean symbols is an accident of birth or education. It is thus unlikely that the human brain has hard-wired neural machinery for recognizing a specific writing system, as literacy is a relatively recent phenomenon and human scripts differ in form significantly from one culture to another. One could question whether the lack of scale invariance found with letters reflects the arbitrary and artificial nature of written text. If so, it would be of interest to examine scale dependence with stimuli that have greater universality and longer evolutionary significance for humans: faces, in particular, may constitute such a class.

We examined the degree of scale dependence for five sets of stimuli that we could characterize in terms of two basic factors: experience and evolutionary significance (Figure 2). (1) Letters constitute a stimulus set with which literate subjects have a high degree of experience, but which as stimuli have low evolutionary significance, for the reasons stated above. (2) Mirror-image letters we considered to represent an intermediate level for experience, in so far as subjects have far less exposure to such letters but can still easily recognize these as transformed letters, and low evolutionary significance set. (3) Novel shapes are those with low experience and low evolutionary significance. (4) Upright faces are as a stimulus set with high evolutionary significance and a high degree of experience, as all humans are raised with significant daily exposure to upright faces. (5) Inverted faces represent high evolutionary significance and low/intermediate degree of experience, in so far as faces tend to be encountered far more frequently in the upright orientation. We used critical band masking (Solomon & Pelli, 1994) to estimate the spatial frequencies that are predominantly used in recognizing these stimuli at various sizes and determined how these critical frequencies change with size.

If scale invariance requires evolutionary time scales to be built into specialized neural mechanisms, then evidence for scale invariance should only be found for the face stimuli. Furthermore, if experience has an impact, then the degree of scale dependence should be less for those stimulus classes with which humans have more experience and familiarity.

Methods

Subjects

Ten subjects (7 females, ages 18–33 years) with normal or corrected-to-normal vision participated in this study. Subjects completed all size conditions of one or more experiments in which they participated, with two exceptions (for details, see Table 1). Because of the lengthy nature of these experiments, not all subjects did all experiments. Rather, each experiment included a different subset of the subjects. Seven subjects participated in the upright faces experiment, five subjects participated in the letters and inverted faces experiments, and four subjects participated in the mirror-image letters and novel shapes experiments. A detailed summary of subject participation in each experiment is shown in Table 1. 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.

Experimental setup

The experimental procedure was implemented on a computer equipped with a Cambridge Research Systems

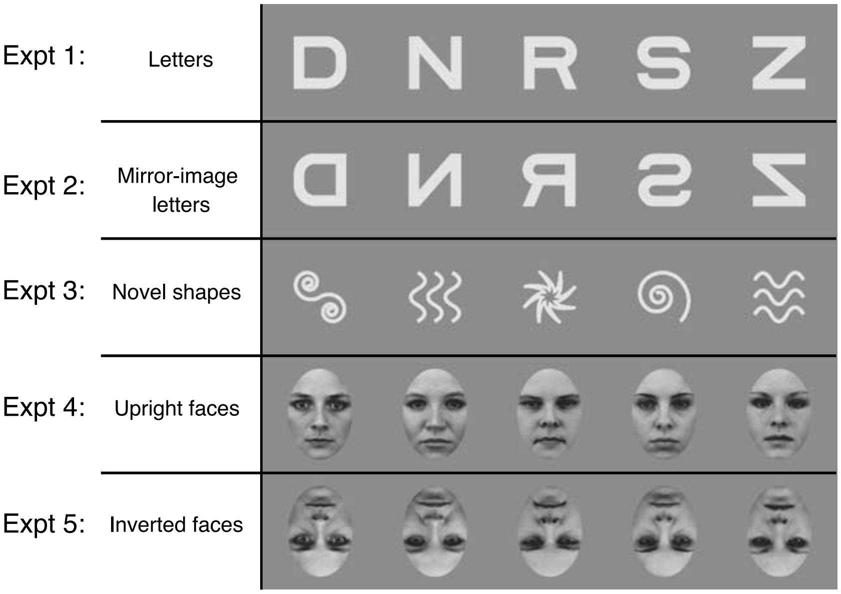


Figure 2. Stimulus sets. All five classes of stimuli are shown. From top to bottom: Experiment 1, letters; Experiment 2, mirror-image letters; Experiment 3, novel shapes; Experiment 4, upright faces; and Experiment 5, inverted faces. Each stimulus class contains five individual stimuli. Discrimination contrast thresholds for each category were measured using a 5-alternative forced-choice paradigm.

	Viewing distance			
Experiments	3×	2×	Standard	Half
Experiment 1: letters	CF, IO*, KD*, KL, LD	–	CF, IO*, KD*, KL, LD	CF, IO*, KD*, KL, LD
Experiment 2: mirror-image letters	BS, CF, IO, KL	–	BS, CF, IO, KL	BS, CF, IO, KL
Experiment 3: novel shapes	BS, CF, IO, KL	–	BS, CF, IO, KL	BS, CF, IO, KL
Experiment 4: upright faces	CF, IO, JDK, KD, SR, NP	CF, IO, JDK, KD, SR, GL	CF, IO, JDK, KD, SR, NP, GL	CF, IO, JDK, KD, SR, NP, GL
Experiment 5: inverted faces	BS, JDK, KD, IO, SR	BS, JDK, KD, IO, SR	BS, JDK, KD, IO, SR	BS, JDK, KD, IO, SR

Table 1. Summary of subject participation. Ten subjects (7 females, ages 18–33 years) participated in the study: BS, CF, GL, IO, JDK, KD, KL, LD, NP, and SR. Each subject participated in one or more of the five experiments. Participants completed all size conditions of the experiment(s) they took part in, with two exceptions: Experiment 4, NP at 2×, and GL at 3×. *Note*: *The Letters data of IO and KD have been published in a previous study (Oruc & Landy, 2009) and reproduced here.

(CRS) VSG 2/3 graphics card and SONY Trinitron 17 in monitor (model GDM-200 PS). The display was gamma corrected using OptiCAL photometer (Model OP200-E) and software provided by CRS. Gamma correction was repeated regularly every month to ensure stable luminance calibration. 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

As stated above, five classes of stimuli were used: (1) letters, (2) mirror-image letters, (3) novel shapes, (4) upright faces, and (5) inverted faces (Figure 2). Each stimulus class contained five individual stimuli, e.g., five letters. All stimuli were 4.7 deg wide at the standard size. Large (double, i.e., 9.5 deg), small (half, i.e., 2.35 deg), and very small (one third, i.e., 1.58 deg) sizes were implemented by displaying the stimuli at half, double, and triple the standard viewing distance. Letters, mirror-image letters, and novel shapes were viewed at three sizes (very small, standard, and large) and upright faces and inverted faces were viewed at four sizes (very small, small, standard, and large). The standard viewing distance was 91.4 cm for letters, mirror-image letters and novel shapes, and 107 cm for upright and inverted faces. All stimuli were grayscale and displayed on a uniform gray background at mean luminance.

Letters, mirror-image letters, and novel shapes

Five letters (D, N, R, S, and Z) in Sloan font (Pelli, Robson, & Wilkins, 1988; available at http://www.psych.nyu.edu/pelli/software.html) were used. These stimuli were the same as the letter stimuli used in a previous study (Oruc & Landy, 2009). Letters represent the *high-experience* set based on reading experience on the order of tens of years. Mirror-image letters were left-right reversed versions of the letter stimuli. Mirror-image letters constitute an *intermediate-experience* set because they are recognized at once and thus share in some of the training for regular letters. However, most observers have had little practice with mirror-image letters. The novel shapes represent the *low-experience* set. To create the novel shape stimuli we designed arbitrary patterns composed of simple strokes on a uniform background. In this respect, the novel shape stimuli were not unlike letters. While all our observers were highly familiar with the letters, they did not have any prior experience with our novel stimulus set and did not receive any experimental training on these before starting the experiment.

We specified Weber contrast for all three types of stimuli (letters, mirror-image letters, novel shapes) as the increment in luminance above the mean luminance, divided by the mean luminance. Stimuli were displayed on a uniform gray background at mean luminance. Stimulus luminance values varied between mean luminance (0% contrast) and maximum luminance (100% contrast). All three types of stimuli were approximately 240 pixels wide, corresponding to 4.7 deg at the viewing distance of 91.4 cm.

Upright and inverted faces

Five female faces displaying a neutral expression were selected from the Karolinska Database of Emotional Faces (Lundqvist & Litton, 1998). Face images were converted to grayscale using Adobe Photoshop CS 8.0 (www.adobe.com). Faces were seen through an oval aperture that was 283 pixels at the widest point, corresponding to 4.7 deg at the viewing distance of 107 cm. The viewing distances for the upright and inverted face stimuli were slightly different

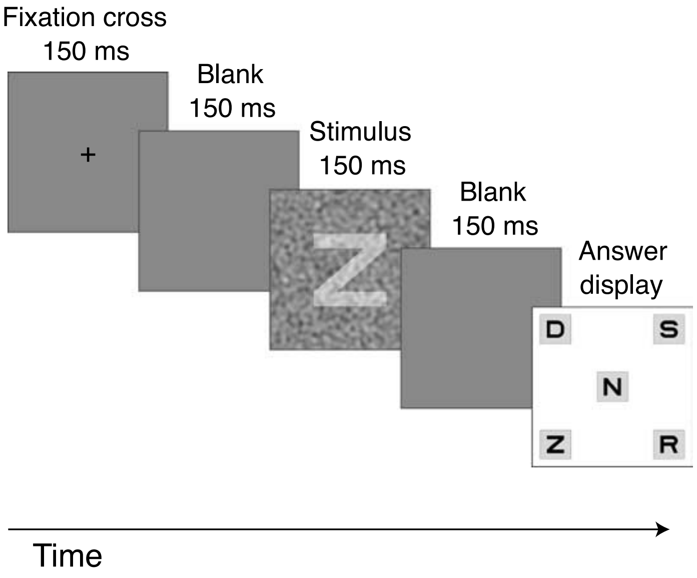


Figure 3. Sequence of events in a typical trial. At each trial one of five possible stimuli (of a fixed stimulus class) is shown for 150 ms, with or without added visual noise. The observer indicates which stimulus they saw by choosing it from a choice display that contains the set of all alternatives. In this example, a typical trial from Experiment 1 is shown where the stimulus class is letters. The procedure is identical for the other stimulus classes/experiments: mirror-image letters, novel shapes, upright faces, and inverted faces.

from that for letters, mirror-image letters, and novel shapes to obtain equal width in visual angles for stimuli with different width in pixels. Faces were aligned spatially within the oval aperture by horizontally centering the tip of the nose and adjusting vertical position to set pupil height to the same level across all faces. Care was taken to ensure that faces chosen lacked obviously distinguishing marks to avoid discrimination based on trivial differences.

We specified root-mean-squared (RMS) contrast for face stimuli, defined as the standard deviation of luminance divided by mean luminance. To ensure standard contrast across all face images prior to experimental manipulation of contrast for threshold measurements, mean luminance was set to 0.5 (half maximum luminance) and RMS contrast inside the oval aperture was normalized to 1. All image manipulation including overlaying an oval mask, horizontal and vertical aligning, and luminance and contrast normalizations were performed using in-house scripts in Matlab (www.mathworks.com).

Noise

There were sixteen noise conditions, including eight low-pass and seven high-pass filtered noises, in addition to a no-noise condition. The noise conditions and the generation of noise masks follow methods used in Oruc and Landy (2009). Noise masks were generated by low- or high-pass filtering Gaussian white noise (40% RMS contrast prior to filtering corresponding to a two-sided noise spectral density of 6.1×10^{-5} deg^2 at 91.4 cm viewing distance) at cutoff frequencies ranging from 0.1 to 17 cpd (at 91.4 cm viewing distance). Each cutoff frequency defined a complementary pair of low-pass and high-pass noise masks that added up to the low-pass noise mask with the highest cutoff. Noise contrast was fixed throughout all five experiments. Low-pass filters were radially symmetrical smooth Butterworth filters defined as

$$B(f) = \frac{1}{(1 + f/f_c)^{10}}, \tag{1}$$

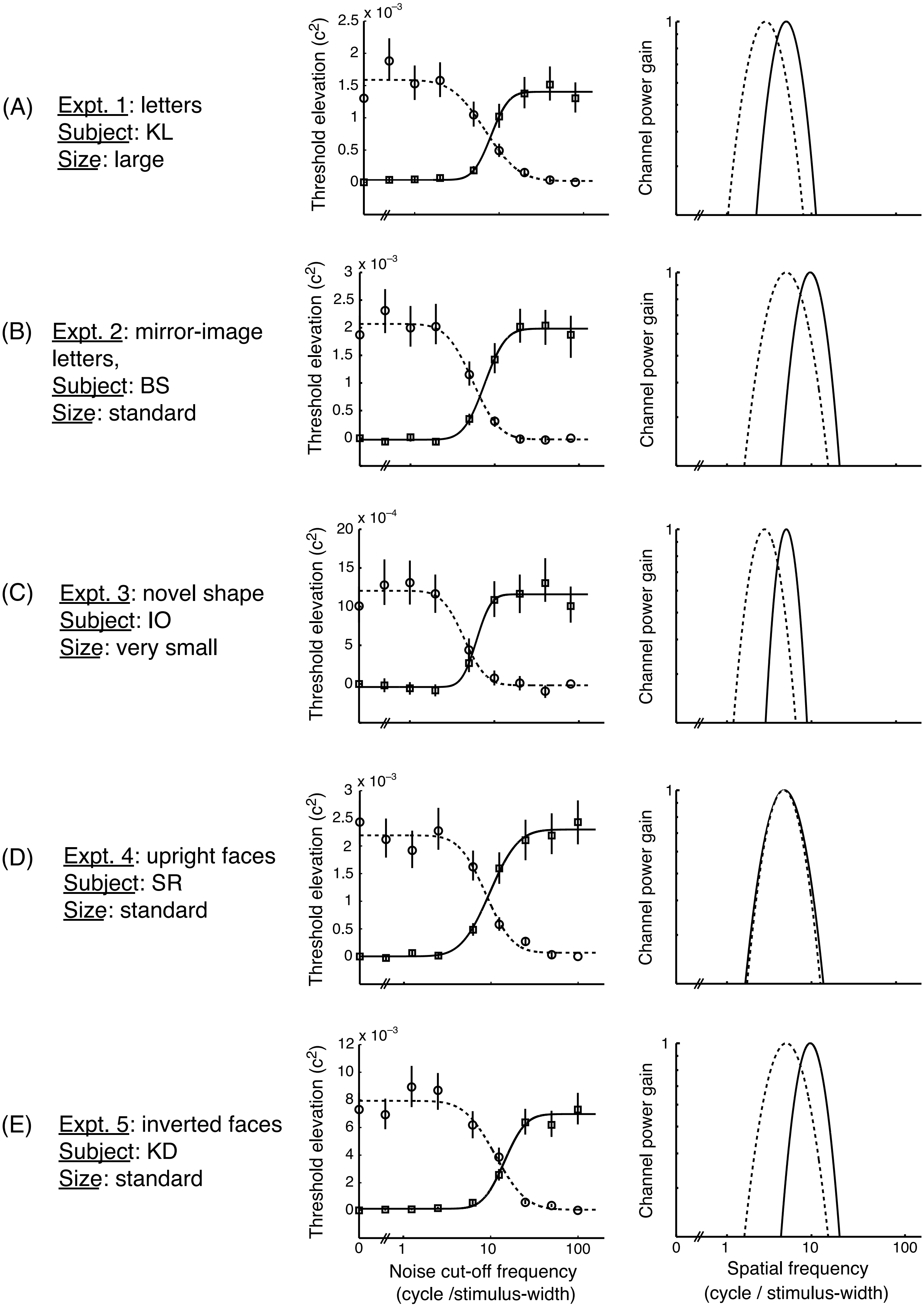
where f denotes frequency, and f_c denotes cutoff frequency. Corresponding high-pass filters were defined as $1 - B(f)$.

We pre-computed noise images that were 16 times larger in area than the actual stimuli (i.e., 4× larger in size in both dimensions). At the start of each block, the appropriate noise image for that block was loaded in the memory. A new noise mask at each trial was obtained by first circularly shifting the pre-computed large noise image by random offsets in both dimensions and then assigning the top left quadrant as the current noise mask. New noise images were pre-computed for every block.

Procedure

We measured contrast thresholds for identification of 1 out of 5 stimuli within a given category embedded in various low- or high-pass filtered noise masks. At each trial, one stimulus from the current category, e.g., the letter Z, out of five alternatives, e.g., D, N, R, S, Z, was shown, with or without added noise according to the trial type. The observer's task was to indicate which one of the five letters they saw. A trial consisted of the following sequence of displays: a 150-ms fixation cross, a 150-ms blank, a 150-ms stimulus display, a 150-ms blank, and finally a choice screen displaying all five alternatives that remained visible until the observer responded (Figure 3). The observer entered their response using keys on the computer keypad that spatially corresponded to the choices screen display. An auditory signal provided feedback indicating whether the response was correct (single beep) or incorrect (double beeps). All five experiments used the identical procedure with different stimulus sets.

The experimental trials were blocked by noise type, resulting in 16 blocks corresponding to the 16 noise conditions (seven high-pass, eight low-pass, and no-noise), completed in a random order. In each block, two estimates of the discrimination threshold for the given noise condition were measured via two randomly interleaved


(A) Expt. 1: letters
Subject: KL
Size: large
(B) Expt. 2: mirror-image letters,
Subject: BS
Size: standard
(C) Expt. 3: novel shape
Subject: IO
Size: very small
(D) Expt. 4: upright faces
Subject: SR
Size: standard
(E) Expt. 5: inverted faces
Subject: KD
Size: standard
Threshold elevation (c^2)
x 10^{-3}
2.5
2
1.5
1
0.5
0
x 10^{-3}
3
2.5
2
1.5
1
0.5
0
x 10^{-4}
20
15
10
5
0
x 10^{-3}
3
2.5
2
1.5
1
0.5
0
x 10^{-3}
12
10
8
6
4
2
0
0
1
10
100
Noise cut-off frequency
(cycle /stimulus-width)
Channel power gain
1
0
1
10
100
Spatial frequency
(cycle / stimulus-width)

staircases, 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 warm-up block started each session to allow the subjects to accommodate to the setting and the experimental procedure.

For letters, mirror-image letters, and novel shapes, subjects completed three size versions (large, standard, and very small) in a random order. For upright face and inverted face experiments, subjects completed four size versions (large, standard, small, and very small) in a random order.

To ensure sufficient familiarity with the stimuli, subjects completed one practice session of 1280 trials (lasting approximately 1 hour) at the standard size prior to the experimental sessions with all stimulus classes except the novel shape stimulus. No practice session was provided for the experiment with novel shape stimulus category to keep exposure minimal prior to experimental data collection.

Data analysis

Contrast thresholds

In each block, two estimates of the contrast threshold in a given noise condition are measured by two randomly interleaved staircases. Observers completed 1–3 blocks for each noise type, resulting in 2–6 threshold estimates. Threshold elevation was computed as the difference between squared threshold contrast (averaged across the 2–6 independent estimates) in a given noise condition and the no-noise condition, which served as the baseline.

Critical frequencies

Threshold elevations increased following a sigmoidal trend for the low-pass noises as the cutoff frequency was increased. As expected, the reverse pattern was evident for the high-pass noise conditions: Threshold elevation was maximum at the lowest cutoff values and gradually decreased to baseline with increasing cutoffs. A cumulative Gaussian was fit to the low-pass threshold elevation data as a function of the logarithm of the noise cutoff frequency. Similarly, a reversed cumulative Gaussian was independently fit to the high-pass threshold elevation data. The derivative of the cumulative Gaussian divided by f, the cutoff frequency, yielded the channel gain. Based on previous results, we expected a moderate degree of off-frequency looking or channel switching (Oruc & Landy, 2009). Therefore, the critical spatial frequency in each condition was estimated by averaging the peak frequencies of the two channel gains independently obtained from the low-pass and high-pass threshold elevation curves. Figure 4 shows example data sets from each one of the five experiments (stimulus categories) to illustrate the data analysis. Data for other subjects and stimulus sizes showed a similar pattern in general and were analyzed in the same way (for example data sets plotted across all sizes for each of the five categories for individual subjects, see Supplementary Figure 1).

Figure 4. Estimation of critical spatial frequency bands: sample data sets from the five experiments. On the left, threshold elevation is plotted as a function of noise cutoff frequency for the low-pass (solid curve) and high-pass (dashed curve) noise masks. Threshold elevation, defined as the difference between squared contrast thresholds at a noise condition and at the no-noise condition (baseline), increases monotonically with noise cutoff frequency for low-pass noise masks (circles) and monotonically decreases with noise cutoff frequency for the high-pass noise masks (squares). Cumulative Gaussians were independently fit to the low-pass and high-pass data. The derivative of the threshold elevation curves divided by cutoff frequency provides an estimate of the power gain. On the right, the two estimates of the power gains obtained independently from the low- and high-pass curves are shown. The peak frequencies estimated from the two curves are averaged to yield the critical spatial frequency at the given condition.

Results

Figure 5 shows the results for all five experiments. We plot critical spatial frequencies as a function of stimulus size. Figures 5A–5E show group data pooled over all subjects in each experiment. The results for letter recognition replicated the data of Majaj et al. (2002), which are shown superimposed on the same graph (Figure 5A). As expected, when letters gets larger, the critical frequency band for letter recognition shifts to higher values in units of *object frequency* (i.e., cycle/letter), characterizing the scale-dependent nature of the process as described before. (If the process was scale invariant, the critical frequency as expressed in units of object frequency would not change with changes in stimulus size.) We also find that both the mirror-image letter (Figure 5B) and novel shape (Figure 5C) results are very similar to that of letter recognition.

On the other hand, the results for upright faces differ from the others and show a distinct bi-phasic pattern (Figure 5D). For small sizes a scale-dependent pattern is evident where critical frequencies (in cycle/face-width) increase with size. However, for larger sizes, the critical frequency curve flattens and displays a relatively scale-invariant pattern in which the critical frequency is fixed in units of object frequency and does not continue to increase further with increasing size. The inverted face data are virtually identical to the upright face data at small sizes but do not show an indication of leveling off at larger sizes. In other words, unlike the upright faces,

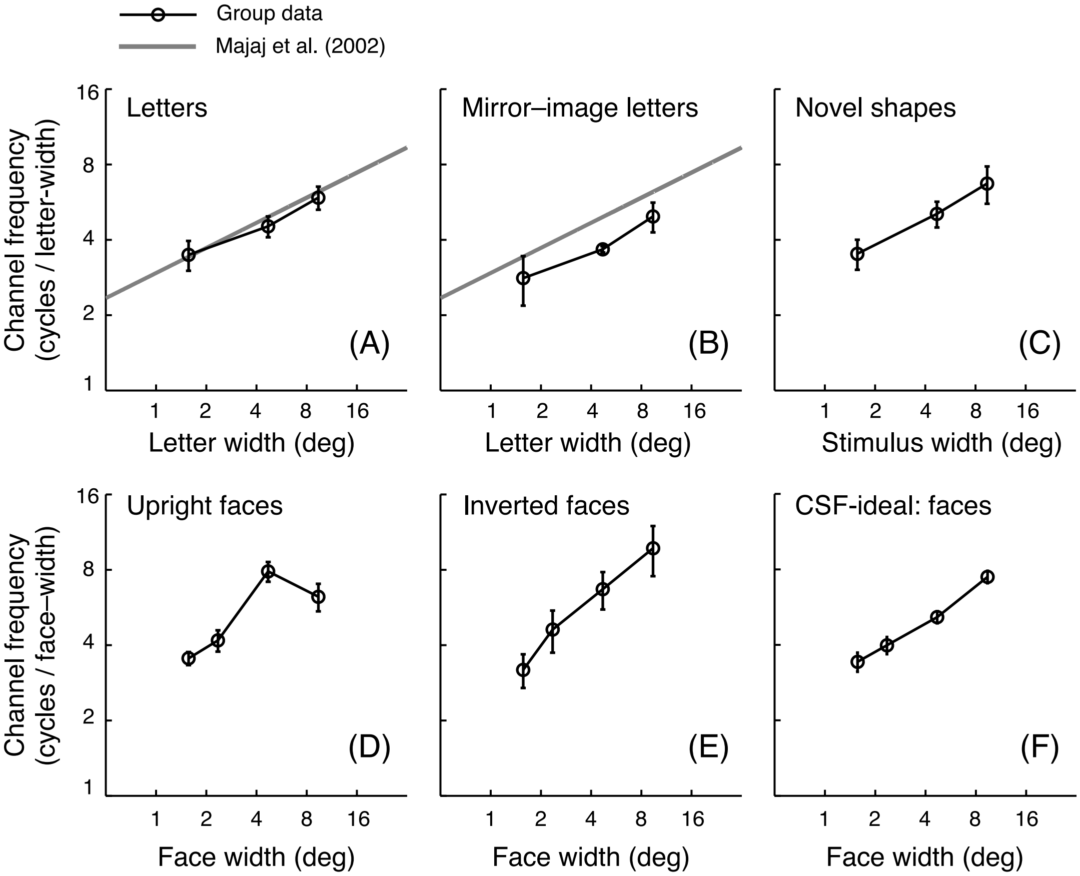


Figure 5. Results. Group data showing critical frequencies used for recognition are plotted as a function of stimulus size, for the human observers for the five stimulus categories (A–E) as well as for the CSF-ideal observer for the face stimuli (F). For the letters and mirror-imaged letters (A–B), results from Majaj et al. (2002) are superimposed for comparison (gray curve).

inverted faces show a single mode with scale dependency, similar to that of the letters.

A closer look at the upright faces data shows that the deflection, or bend, that is apparent in the group data for larger faces (Figure 5D) is in fact consistent across all seven subjects, plotted separately in Figure 6A. A Kruskal–Wallis one-way ANOVA showed that main effect of size on channel frequency was significant for upright faces ($p < 0.05$). Pairwise comparisons using Wilcoxon signed-rank tests showed significant differences between all size pairs (all p's < 0.05, one tailed) with the exception of standard and large size ($p > 0.2$), consistent with the observed pattern that the critical frequencies increase with size at first, but beyond a certain size this trend of increase fades.

To further examine whether this pattern shift represents a flattening effect or a drop in channel frequency with larger sizes, we had one subject sample the sizes more densely and complete the experiment at nine sizes (compared to the four in our original design). This data set, shown in Figure 6B, is fit well by a cumulative Gaussian and thus confirms that the upright face recognition is characterized by two distinct phases, one for small faces showing a scale-dependent pattern in which channel frequencies increase with size, and one for larger faces, showing a scale-invariant pattern in which channel frequency is constant, independent of size.

We have also looked at the width of the channel gains across our five stimulus classes. We found that mean width (full bandwidth at half height) overall was 1.69 ± 0.83 octaves (mean ± *SD*) consistent with former reports for letters (Chung et al., 2002; Majaj et al., 2002; Oruc & Landy, 2009) with similar bandwidths for the different stimulus classes (faces: 1.61 ± 0.82, inverted faces: 1.88 ± 0.89, letters: 1.62 ± 0.84, mirror-image letters: 1.77 ± 0.82, novel shapes: 1.53 ± 0.76). Overall, there was no indication that stimulus size or category affected bandwidths in a systematic way (see Supplementary Figure 2).

Discussion

We have characterized our five stimulus classes based on two factors: evolutionary significance and amount of experience. We find that recognition of object classes with minimal evolutionary significance, i.e., letters, mirror-image letters, and novel shapes, show a clear scale-dependent

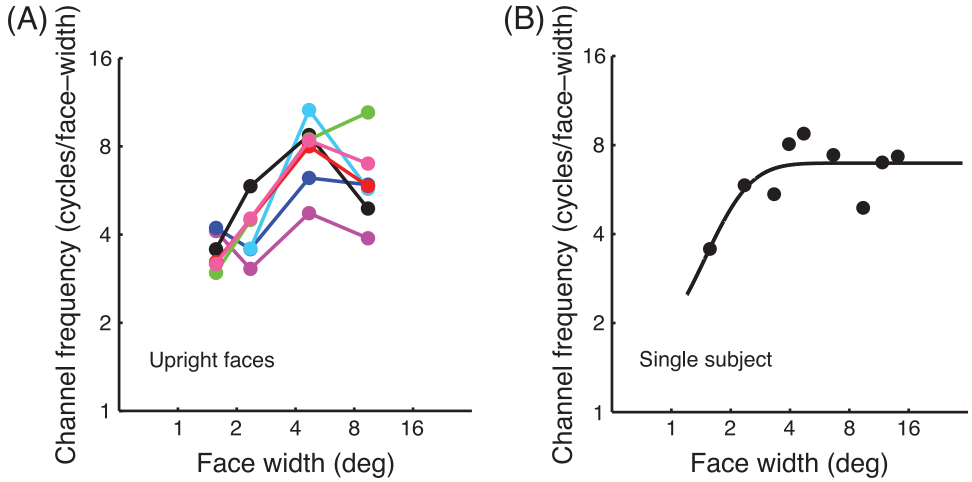


Figure 6. Upright face data. (A) Data for all seven subjects who participated in the upright faces experiment are shown. The characteristic bend, or deflection, apparent in the group data (Figure 5D) for larger faces is also evident at the individual subject level. (B) Data for one subject who completed the upright faces experiment at nine face sizes are shown. This set of data is well fit by a cumulative Gaussian, suggesting that recognition of upright faces is characterized by two distinct regimes: a scale-dependent regime for small sizes and a scale-invariant regime for larger sizes.

pattern of processing across the range of sizes tested. The data for letters, the *high-experience* set, replicates the results of previous studies (Majaj et al., 2002; Oruc & Landy, 2009) closely. The results for mirror-image letters and novel shapes, the *intermediate-experience* and *low-experience* sets, respectively, also follow the pattern found with letters, suggesting that, for stimuli with minimal ecological significance, experience in the course of one's lifetime has little impact on the degree of scale dependence.

Upright faces were the only stimulus category we tested that showed signs of scale invariance in our study, but this was limited to relatively larger face sizes. Our group data showed that, when faces were 4.7 degrees of visual angle in width or less, they showed scale dependency in their processing, just like other objects. Based on a median face width of 12.8–14 cm (Poston, 2000), 4.7 degrees corresponds to viewing a real face from approximately 1.6 m away, which falls in the middle of the range what Hall (1966) termed *close-phase social distance* (4–7 feet, i.e., between 1.2 and 2.1 m), characterizing the distance people use when working together or in a casual social gathering. Thus, when faces are viewed at sizes typically encountered during social interactions, they show signs of scale invariance and are processed in a qualitatively different manner than when they are smaller and hence more distant.

The inverted face data (Figure 5E) are virtually identical to the upright face data (Figure 5D) with one crucial difference: there is no evidence for the flattening of the critical frequency curve with increasing size from 4.7 to 9.5 degrees of visual angle in width. These results are consistent with those of Gaspar, Sekuler, and Bennet (2008) who found that critical frequencies for recognizing upright and inverted faces did not differ for faces with widths of 2.3 degrees. Our data show the same pattern at that size. In fact, considering that both upright and inverted versions of a particular face stimulus have the identical spatial frequency power spectrum, i.e., identical information content, it should not be surprising that the frequencies critical for recognition would be the same in each case. The unexpected aspect of the data is the divergence of the results at the larger face size. The fact that the scale invariance at these sizes is seen only for upright faces suggests that experience has an impact on scale dependency but only for highly ecologically significant stimuli.

One possible explanation for the bend observed for the upright face stimuli is that the critical frequency curve cannot continue moving up to higher frequencies with larger sizes because these are too high to be visible or that visibility is sufficiently impaired that it impacts the usefulness of the diagnostic information. This is unlikely given the fact that this deflection does not occur for inverted faces: information at these higher spatial frequencies must be visible and useful for recognition. Nevertheless, to conclusively exclude reduced contrast sensitivity and lack of diagnostic information at the higher spatial frequencies as a potential explanation of the characteristic bend, we ran an ideal observer simulation and compared human data to that of the model. The ideal observer is a computer model that goes through the same experimental tasks as the human observers and is able to use all information available in an optimal fashion. As such, it serves as a benchmark of the best possible performance. Under normal circumstances, the ideal observer's behavior is scale invariant by definition. In this particular case, though, we furnished the model with the contrast sensitivity profiles of our human observers such that the model was also hampered by the same resolution and sensitivity constraints as the human observers.

To accomplish this, we first measured contrast sensitivity functions of four individual observers who participated at all size conditions of the upright faces experiment. We then computed for each individual observer the equivalent input noise, i.e., the noise spectrum at which an ideal observer would have the same contrast sensitivity profile as the human observer. Our CSF-ideal observer then went through the same experimental trials as the humans in which the stimuli were corrupted by one of the 16 noise masks of the given condition as well as the equivalent input noise of the individual observer. Thus, we produced CSF-ideal model predictions for each observer separately.

At each trial, the CSF-ideal observer was shown a noisy stimulus and asked to chose which one of the five faces the stimulus corresponded to. The CSF ideal had knowledge of the five face templates, the contrast at the current trial, and the statistics of the two noise masks. The response of the CSF ideal was based on maximum likelihood given the noisy stimulus. Further details of the CSF-ideal observer as well as the human contrast threshold measurements can be found in Oruc and Landy (2009, Appendices A and B).

The results of this simulation are shown in Figure 5F. We find that similar to the human data on inverted faces, the CSF ideal uses the higher portion of the face spatial frequency spectrum at larger sizes, with no sign of a bend. If these higher frequencies were not available to the human observers due to contrast sensitivity constraints, then they would also not be available to the CSF-ideal observer. This result confirms that the characteristic deflection pattern seen in the human upright face data is not due to visibility of the image contents but likely represents a shift in strategy for *how* information is used by the human observers to accomplish recognition at different sizes.

Work by Sinha and colleagues demonstrate the remarkable ability of human observers to recognize faces in very low-resolution images (Sinha, 2002a, 2002b; Sinha, Balas, Ostrovsky, & Russell, 2006; Yip & Sinha, 2002). This result is consistent with our current findings: when faces are viewed at the smallest size, the critical frequencies used to recognize faces correspond to approximately 2.3 cpd. In other words, at our smallest size, face images would be highly recognizable at a 4 × 5 pixel image resolution, very similar to the results of Sinha et al. Our present findings also extend a prediction: we argue that such drastically low-resolution face images should become less recognizable and require higher image resolution when viewed at larger sizes, e.g., at 5 degrees face width or larger.

A scale-dependent relationship between critical spatial frequencies and stimulus size is consistently found for various object categories including letters (Chung et al., 2002; Majaj et al., 2002; Oruc & Landy, 2009), words (Chung & Tjan, 2009), mirror-imaged letters, novel shapes, inverted faces, and small upright faces. We show that, so far, the only exception occurs for upright faces viewed at distances typical of social interactions. Our interpretation of this result is based on the special evolutionary status of faces as a stimulus class, though there are other ways in which our stimulus sets differ. For one, although our images are all two-dimensional, faces are complex three-dimensional shapes containing a gradation of gray-scale tones whereas the letters and other similar stimuli are simple two-tone images of two-dimensional objects, consisting of strokes on a uniform background. In addition, while letters differ by the varying spatial configuration of strokes, all faces share the same configuration and are differentiated by subtle differences in that common template. However, these image- and object-based differences do not explain the difference we observe between upright and inverted faces or between small and large upright faces. A second difference between faces and letters lies in the neuroanatomy of their processing: while functional imaging and lesion studies show activity in both hemispheres related to both words/letters and faces, the activity in the right hemisphere dominates for faces whereas that in the left hemisphere dominates for letters (Cohen et al., 2002; Kanwisher, McDermott, & Chun, 1997). At this point, it remains possible that this neuroanatomic difference is as relevant as the ecological difference between letters and faces to the divergent results for these stimuli.

At ecologically relevant sizes, the spatial frequencies used for recognition are lower than what would be expected based on the extrapolation of the scale-dependent component of the upright face data at smaller sizes (Figure 5D) and also based on a comparison to the inverted face data (Figure 5E) and the CSF-ideal observer results (Figure 5F). What does this switch in the usage of spatial frequency signify? Given the fact that this deflection occurs for upright faces but not inverted, it is conceivable that it represents a switch to a recognition strategy based on holistic, or configural, processes (Maurer, Grand, & Mondloch, 2002; Sergent, 1984; Tanaka & Farah, 1993). It has been argued before that holistic processes may depend on lower spatial frequencies (Goffaux & Rossion, 2006), as opposed to generic objects or inverted faces, which are recognized via part-based processes that may rely on higher spatial frequencies.[1] Our present findings, which show that at ecologically relevant sizes faces are recognized using spatial frequencies that are *relatively lower* than expected (Figures 5D–5F), may indicate that at these sizes a holistic strategy is favored whereas at smaller sizes a part-based strategy is used. This, however, must be considered merely a speculation at present. Nevertheless, these findings may support arguments that faces are special and that their recognition may involve specialized expert strategies, but perhaps mainly at ecologically relevant sizes.

Conclusions

A large literature exists now that examines the critical spatial frequencies used for recognition of various visual stimuli such as letters and faces. Most of these studies measured critical spatial frequencies at a single stimulus size (e.g., Costen, Parker, & Craw, 1996; Gaspar et al., 2008; Gold, Bennett, & Sekuler, 1999; Näsänen, 1999; Peli, Lee, Trempe, & Buzney, 1994; Scharff, Hill, & Ahumada, 2000). Whenever critical frequencies were measured as a function of size (Chung et al., 2002; Chung & Tjan, 2009; Majaj et al., 2002; Oruc & Landy, 2009), a scale-dependent pattern of results was observed, despite a prevalent intuition that object recognition should be minimally influenced by object size. Scale dependence holds true for a large variety of stimulus classes tested (Chung et al., 2002; Chung & Tjan, 2009; Majaj et al., 2002; Oruc & Landy, 2009), with our results showing one important exception: upright faces viewed at sizes with ecological and social significance.

Acknowledgments

This work was supported by CIHR MOP-77615. IO was supported by Natural Sciences and Engineering Research Council, Discovery Grant RGPIN 355879-08. JB was supported by a Canada Research Chair and Michael Smith Foundation for Health Research Senior Scholar Award.

Commercial relationships: none.
Corresponding author: İpek Oruç.
Email: ipor@interchange.ubc.ca.
Address: Human Vision and Eye Movement Laboratory, VGH Eye Care Center, 2550 Willow Street, Vancouver, BC, V5Z 3N9, Canada.

Footnote

[1] Our results also provide a means to reconcile these data and those that claim no difference between the spatial frequencies used to recognize upright and inverted faces (Gaspar et al., 2008; Goffaux & Rossion, 2006). Gaspar et al. (2008) measured critical frequencies at the size of 2.3 degrees per face width, at which we find a similar result, whereas Goffaux and Rossion (2006) used larger faces (4.1 degrees), which may explain the difference they observed.

References

Ahumada, A. J., Jr., & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A, 2,* 1133–1139.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 121–133.

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

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

Chung, S. T. L., & 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]

Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain, 125,* 1054–1069.

Costen, N., Parker, D., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58,* 602–612.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision Research, 40,* 473–484.

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

Goffaux, V., & Rossion, B. (2006). Faces are “spatial”-holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1023–1039.

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,* 3537–3560.

Hall, E. T. (1966). *The hidden dimension* (1st ed.). New York: Doubleday & Co.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research, 25,* 239–252.

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,* 1165–1184.

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

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39,* 3824–3833.

Oliva, A., Torralba, A., & Schyns, P. G. (2006). Hybrid images. *ACM Transactions on Graphics, 25,* 527–532.

Oruc, I. (2003). *Three studies on perception of texture-defined form and depth cue combination*. Unpublished doctoral dissertation, New York University, New York.

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,* 1493–1506.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31,* 1399–1415.

Peli, E., Lee, E., Trempe, C. L., & Buzney, S. (1994). Image enhancement for the visually impaired: The effects of enhancement on face recognition. *Journal of the Optical Society of America A, Optics, Images Science, and Vision, 11,* 1929–1939.

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

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

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences, 2,* 187–199.

Poston, A. (2000). *Human engineering design data digest*. Retrieved from http://hfetag.com/hfs_docs.html.

Scharff, L. F., Hill, A. L., & Ahumada, A. J., Jr. (2000). Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express, 6,* 81–91.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75,* 221–242.

Sinha, P. (2002a). Identifying perceptually significant features for recognizing faces. *Proceedings of SPIE, 4662,* 12–21.

Sinha, P. (2002b). Recognizing complex patterns. *Nature Neuroscience, 5,* 1093–1097.

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE, 94,* 1948–1962.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369,* 395–397.

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

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

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 31,* 995–1003.