The human visual system is optimised for processing the spatial information in natural visual images

C.A. Párraga*, T. Troscianko* and D.J. Tolhurst†

A fundamental tenet of visual science is that the detailed properties of visual systems are not capricious accidents, but are closely matched by evolution and neonatal experience to the environments and lifestyles in which those visual systems must work [1–5]. This has been shown most convincingly for fish [6] and insects [7]. For mammalian vision, however, this tenet is based more upon theoretical arguments [8–11] than upon direct observations [12,13]. Here, we describe experiments that require human observers to discriminate between pictures of slightly different faces or objects. These are produced by a morphing technique that allows small, quantifiable changes to be made in the stimulus images. The independent variable is designed to give increasing deviation from natural visual scenes, and is a measure of the Fourier composition of the image (its second-order statistics). Performance in these tests was best when the pictures had natural second-order spatial statistics, and degraded when the images were made less natural. Furthermore, performance can be explained with a simple model of contrast coding, based upon the properties of simple cells [14–17] in the mammalian visual cortex. The findings thus provide direct empirical support for the notion that human spatial vision is optimised to the second-order statistics of the optical environment.

Addresses: *Department of Experimental Psychology, Bristol University, 8 Woodland Road, Bristol BS8 1TN, UK. †Department of Physiology, Cambridge University, Downing Street, Cambridge CB2 3EG, UK.

Correspondence: C.A. Párraga
E-mail: Alej.Parraga@bris.ac.uk

Received: **16 July 1999**
Revised: **14 October 1999**
Accepted: **23 November 1999**

Published: **17 December 1999**

Current Biology 2000, 10:35–38

0960-9822/00/$ – see front matter

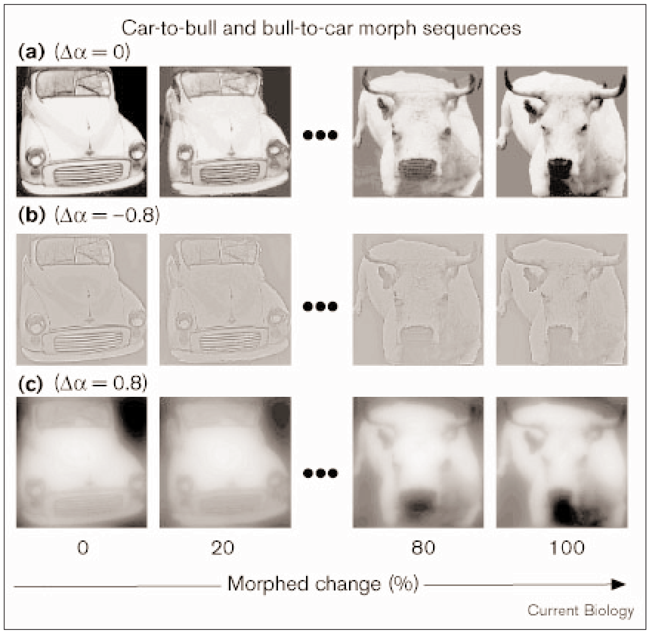
Results and discussion

In order to prove that visual coding is optimised for working in the natural world, it is necessary to demonstrate experimentally that an animal really does ‘see’ natural things better than it does unnatural ones. As much of our visual processing of static or slowly changing images involves the discrimination of form, we have developed a naturalistic form-discrimination task as a likely exemplar of the kind of task for which our visual system might have evolved. This task requires a human observer to distinguish between subtly different monochrome pictures in which the component objects differ slightly in shape, position, texture and brightness (Figure 1a). The observer must discriminate visually between pictures of, say, very similar faces or very similar objects. We asked whether the observer’s ability to perform this exemplar task would be degraded if the spatial contrast of the pictures is made unnatural in some way. Photographs of natural scenes have remarkably similar second-order statistics, which are summarised succinctly by noting that their amplitude (or Fourier) spectra are approximated quite well by the single, simple equation:

$$Amplitude(f) \propto f^{-\alpha} \qquad (1)$$

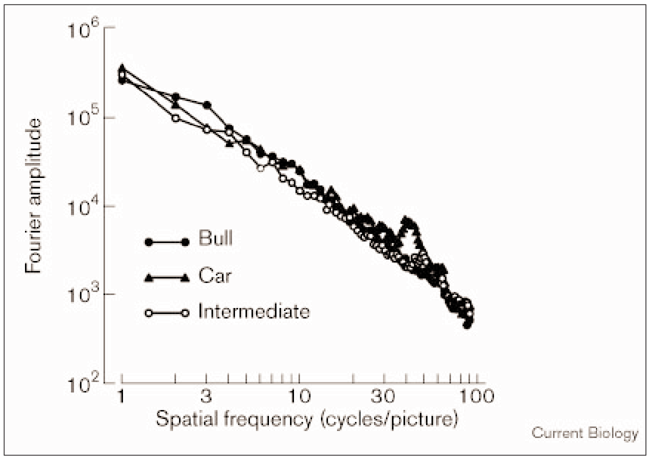
Where f is spatial frequency and α (the slope constant) ranges from about 0.7 to about 1.6 [18–20] (Figure 2). Thus, we can make a set of pictures more or less unnatural in a systematic way by decreasing (Figure 1b) or increasing (Figure 1c) the slopes of their amplitude

Figure 1



Examples of pictures from two of the four morph series used in our psychophysical experiments. **(a)** Parts from the original car-to-bull morph sequence (left to right) in which all pictures have approximately ‘natural’ statistics. The bull-to-car sequence runs right to left. **(b,c)** The same pictures after they have been processed to (b) decrease (‘whiten’) and (c) increase (‘blur’) the slopes of the amplitude spectra. Slope offset ($\Delta\alpha$) for each sequence is shown.

Figure 2



Current Biology

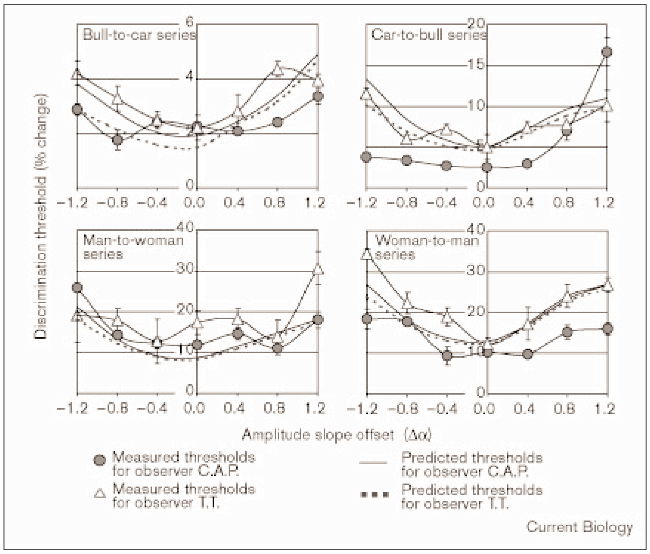
Plots of the spectral amplitude versus spatial frequency for the Fourier spectrum of the car (0% change), the bull (100% change) and an intermediate morphed picture from Figure 1a. The plot shows the similarity between the spectra and, hence, the second-order statistics of all the pictures. All slopes (α in Equation 1) fall within a close range (~1.5).

spectra. If performance is optimal when the second-order statistics correspond to those of natural scenes, we will be able to conclude that those statistics have been important in shaping the function of our visual system.

Figure 3 shows the experimental results for two observers, C.A.P. and T.T., who are authors of this paper. The main effects were also replicated on a psychophysically experienced but naive observer. Each observer was presented with four types of reference stimulus: pictures of a bull, a car, a man's face and a woman's face. The bull and the car formed one morph sequence; the man's face and the woman's face formed another. The observer had to identify test pictures that were different from the reference, in the direction of the other end of the morph series (Figure 1). Thus, a car would become slightly more like the bull, the bull more like the car, the man's face more like the woman's, and the woman's face more like the man's.

The discrimination thresholds in Figure 3 are expressed as percentage movement through the morph series. The plotted points show the discrimination thresholds and the error bars show the standard errors estimated by the probit fitting procedure. The higher the number, the worse the performance. The abscissae show the deviation from the 'normal' value of α, the spectral slope (Equation 1). A positive amplitude slope offset means that the slope was steeper than normal: the scene was 'blurred' (Figure 1c). A negative amplitude slope offset means that the slope was shallower: the scene was 'whitened' (Figure 1b). Zero offset means that the picture had the second-order statistics of natural scenes.

Figure 3



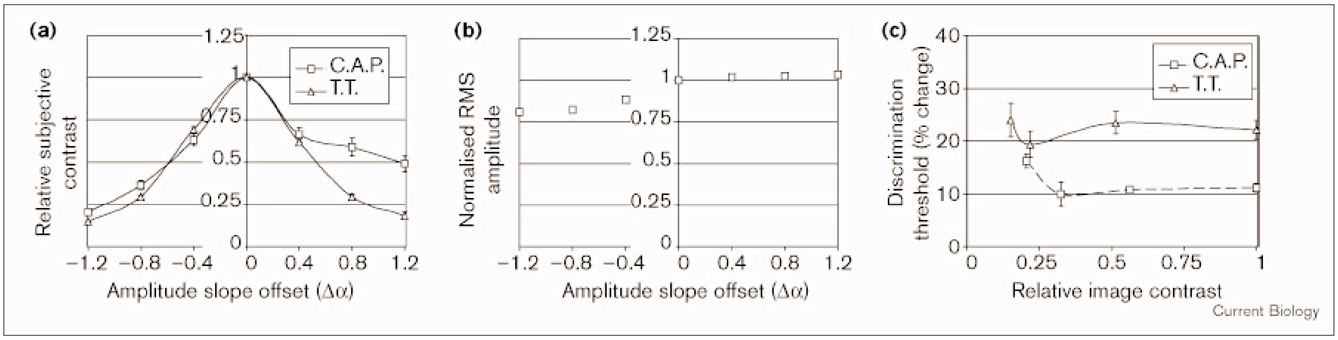
Current Biology

Experimental results obtained for observers C.A.P. and T.T. The plots show the discrimination thresholds for all four morph sequences versus the deviation (Δα) from the natural value of α. Threshold is expressed as a percentage movement through the morph continuum, with 0% representing one of the two original pictures used to create the morph sequence and 100% representing the other original picture. The error bars show the standard error. The curves show the predictions made by our model of local-contrast discrimination.

Our prediction is that, if natural scenes are optimally encoded, the thresholds at zero amplitude slope offset will be lower than the other values. And, indeed, the eight experimental sets do generally show that threshold is lowest between amplitude slope offsets (Δα) of –0.4 and +0.4, and that threshold rises more or less symmetrically from this lowest point. We can show this more formally by fitting two linear equations to each data set and looking at where they intersect, or by fitting second-order polynomials. The mean intersection point for pairs of lines was at a slope offset (Δα) of –0.071 (standard deviation (SD) 0.425). The mean slope offset of the minima of the fitted polynomials was $\Delta\alpha = -0.016$ (SD = 0.175). The χ^2 of these fits were compared to the χ^2 resulting from fitting a single line; on average, they were 13 and 6 times smaller, respectively.

The solid and dashed curves in Figure 3 are predictions of the thresholds for each observer and for each morphed sequence, made by a model of local contrast discrimination between the reference picture and test pictures that progressively differ from it in the morph series. This model has as its input each observer's contrast-sensitivity function (measured with sinusoidal gratings for this purpose). It locally decomposes each scene into its constituent spatial-frequency components, and measures the energy in spatial-frequency bands of amplitude 1.5 octaves [15–17]. It then compares the energy of each filter

Figure 4



Contrast variations across images. **(a)** The relative contrast of pictures of the woman's face when filtered has different amplitude slope offsets. The observers reduced the contrast of the picture with zero offset (which has unity relative contrast) until it appeared to match the contrast of the pictures with other slope offset. **(b)** The normalised root mean square (RMS) amplitude (square root of power) of the same pictures is plotted on the same scale as in (a); note that subjective contrast (a) is not closely related to physical power. **(c)** Results for two observers, who measured their thresholds for discriminating pictures in the woman-to-man series, with pictures presented at a variety of attenuations. The lowest contrasts in this experiment matched the subjective contrasts of picture with slope offsets of ±1.2.

in the two images, and computes whether this lies above the discrimination threshold as measured by the standard contrast discrimination function of, for example, Legge and Foley [21]. The predicted threshold is obtained by computing that morphed image which has just suprathreshold discriminability in at least one spatial-frequency band. The fits of these predicted curves to the data are reasonably good, particularly when considering that the inputs to the model are only the detectability and discriminability of sinusoidal gratings whereas, in this task, observers were discriminating complex images. There are no free parameters to the model, and the curves have not been shifted up or down the ordinate to enhance the fit.

It is noticeable in Figure 1 that changing the slopes of the amplitude spectra changes the apparent contrast of the pictures. Figure 4a shows that the subjective contrast of the picture of the woman is greater for the natural slope than for either steeper or shallower slopes. This is different from physical measurements of image content (Figure 4b), which do not peak at the natural slope. The effects of reducing the contrast of the original pictures without any change in the amplitude slope were investigated. Figure 4c shows that for two observers, the thresholds for discriminating between morphed pictures are little affected by modest changes in the actual contrast of the pictures. This implies that the optimisation of performance for discriminating morphed images with a natural spectral slope is not a trivial result of differences in contrast.

We made our morphed images unnatural by making their second-order statistics (their power spectra) abnormal because this image manipulation was especially straightforward. There are, of course, other image manipulations that would make images unnatural. For instance, the phase spectra of natural images are said to be more crucial than the power spectra to the appearance of pictures [22]. Phase spectra might reflect the third- or higher-orders of image statistics [23], and we would expect that changes to the phase spectra would make our morphed sequences even harder to distinguish than do the changes in power spectra that we report.

Our results indicate that our visual system does perform optimally when the second-order statistics of the presented pictures are natural. Furthermore, a simple model of contrast discrimination does a good job of predicting the data. This optimisation is also manifest by the finding that the natural images in our experiments had higher subjective contrast than those with unnatural spectral slopes, consistent with the proposal that the amplitude spectra of natural scenes might be most appropriately sampled by neurons with natural bandwidths [18,19].

Materials and methods

Visual stimuli

Two different morphed picture sequences were created from monochrome digitised pictures (128 × 128 pixels; 256 grey levels). In the first sequence, the face of a man slowly changes into the face of the woman (the man-to-woman sequence). Here, the shape, contrast and texture vary 2.5% from each picture to the next [24,25]. The second consists of a morph between a picture of a car and a picture of a bull (car-to-bull sequence, Figure 1a). Here, an effort was made to match the salient features of the front of the car (lamps, radiator, the sides of the windscreen) to the salient features of the bull's face (eyes, nose, horns). Given the greater difference between both original images, observers were able to discriminate much smaller percentages of change, and so smaller morph steps (0.5%) were needed. Pictures were presented on a Sony Trinitron monitor driven by a Cambridge Research Systems VSG 2/4 Graphics Card, which was able to compensate for luminance nonlinearities in the display [26]. The pictures measured 8.5 × 8.5 cm in the centre of the display and, since the observer sat 2 m from the display, they subtended 2.43° square at the observer's eyes. Each pixel subtended an angle of approximately one arc min. All other parts of the screen (36 × 29.5 cm), which were not occupied by the stimulus, had a fixed luminance of 85 cd/m^2, the midpoint in the luminance range of the pictures. The observers were allowed to fixate freely.

From the initial sequence of morphed pictures (Figure 1a), further sets were made in which the spectral slope of the component pictures was increased or decreased from the natural value. The Fourier transform of each image was taken and the amplitude spectrum was multiplied by a filter of the form:

$$Weight(f) \propto f^{-\Delta\alpha} \tag{2}$$

where f is spatial frequency and $\Delta\alpha$ determines by how much the amplitude spectrum is made steeper (positive values, Figure 1c) or shallower (negative values, Figure 1b). The new images were constructed by inverse Fourier transformation. The pixel values had to be scaled to fit within the display limit of 127 grey levels on each side of the mean luminance. In scaling, a set of images at any one spectral slope, a single scale factor was used for all images in the set so as to maintain any relative differences in spectral power. To avoid spurious cues resulting from edge effects in the Fourier transform, the pictures were smoothed at the edges with a gaussian roll-off (SD = 15 pixels).

Experiments

Thresholds were measured using a modified two-alternative forced choice paradigm. The observer was presented with three pictures sequentially; each was presented for 500 msec with intervals of 200 msec between. The second presentation always contained a copy of the reference picture; one of the other two presentations (chosen randomly by computer) contained an identical copy of the reference, while the remaining presentation contained a morphed picture. The observer had to press mouse buttons to tell the controlling computer whether the morphed picture was the first or the last of the sequence of three. Auditory feedback was given as to whether the choice was correct. The discrimination task was made harder or easier using a conventional staircase technique, in order to find how much morphing was required to just allow discrimination. If the observer correctly identified the presentation containing the morphed image five times, then a new morphed image was chosen that would be harder to discriminate from the reference. If, however, the observer made one or more errors in a sequence of five trials, then an easier morphed image was chosen for subsequent trials. In a single experiment, four different spectral slopes were randomly chosen and the thresholds for these were measured concurrently. Two independent staircases were run for each slope.

Data analysis

The overall results from 200 trials on each spectral slope were plotted as psychometric functions which were fitted with cumulative normal curves. Threshold was taken as the percentage of morphing that would allow the observer to correctly identify the interval containing the morphed stimulus on 74% of trials. The fitting procedure allowed an estimate of the standard error of the estimated threshold [27].

Acknowledgements.

We are grateful to P.J. Benson for preparing the man-to-woman morph sequences for us, and to I.D. Gilchrist for acting as an observer. C.A.P. and the project as a whole were supported by the MRC (Grant No G9706380).

References

1. Barlow HB: **Possible principles underlying the transformation of sensory messages.** In *Sensory Communications*. Edited by Rosenblith WA. Cambridge, Massachusettes: MIT Press; 1961:217-234.
2. Marr D: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: WH Freeman; 1982.
3. Laughlin SB: **Matching coding to scenes to enhance efficiency.** In *Physical and Biological Processing of Images: Proceedings of an International Symposium*. Edited by Braddick OJ, Sleigh AC, Rank Prize Funds. Berlin; New York: Springer-Verlag; 1983:42-52.
4. Srinivasan MV, Laughlin SB, Dubs A: **Predictive coding – a fresh view of inhibition in the retina.** *Proc Roy Soc Lond B* 1982, **216:**427-459.
5. Van Hateren JH: **Theoretical predictions of spatiotemporal receptive-fields of fly LMCs, and experimental validation.** *J Comp Phys* 1992, **171:**157-170.
6. Lythgoe JN: **Evolution of visual behaviour.** In *Evolution of the Eye and Visual System.* Edited by Cronly-Dillon J, Gregory R. London: Macmillan Press; 1991:3-14.
7. Weckstrom M, Laughlin SB: **Visual ecology and voltage-gated ion channels in insect photoreceptors.** *Trends Neurosci* 1995, **18:**17-21.
8. Hancock PJB, Baddeley RJ, Smith LS: **The principal components of natural images.** *Network Comp Neural Syst* 1992, **3:**61-70.
9. Atick JJ, Redlich AN: **What does the retina know about natural scenes?** *Neural Comp* 1992, **4:**196-210.
10. Olshausen BA, Field DJ: **Emergence of simple-cell receptive field properties by learning a sparse code for natural images.** *Nature* 1996, **381:**607-609.
11. Van Hateren JH, Van Der Schaaf A: **Independent component filters of natural images compared with simple cells in primary visual cortex.** *Proc Roy Soc Lond B* 1998, **265:**359-366.
12. Dan Y, Atick JJ, Reid RC: **Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory.** *J Neurosci* 1996, **16:**3351-3362.
13. Tolhurst DJ, Laughlin SB, Lauritzen JS: **Modelling the variation in the contrast of natural scenes, and the function of contrast normalization in the mammalian visual cortex.** *J Physiol Lond* 1997, **504:**123-124.
14. Hubel DG, Weisel TN: **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.** *J Physiol Lond* 1962, **160:**106-154.
15. Movshon JA, Thompson ID, Tolhurst DJ: **Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex.** *J Physiol Lond* 1978, **283:**101-120.
16. Tolhurst DJ, Thompson ID: **On the variety of spatial-frequency selectivities shown by neurons in area-17 of the cat.** *Proc Roy Soc Lond B* 1981, **213:**183-199.
17. DeValois RL, Albrecht DG, Thorell LG: **Spatial-frequency selectivity of cells in macaque visual cortex.** *Vision Res* 1982, **22:**545-559.
18. Burton GJ, Moorhead IR: **Color and spatial structure in natural scenes.** *Appl Optics* 1987, **26:**157-170.
19. Field DJ: **Relations between the statistics of natural scenes and the response properties of cortical-cells.** *J Opt Soc Am* 1987, **4:**2379-2394.
20. Tolhurst DJ, Tadmor Y, Chao T: **Amplitude spectra of natural images.** *Ophthalm Physiol Opt* 1992, **12:**229-232.
21. Legge GE, Foley JM: **Contrast masking in human vision.** *J Opt Soc Am* 1980, **70:**1456-1471.
22. Piotrowski LN, Campbell FW: **A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase.** *Perception* 1982, **11:**337-346.
23. Thomson MGA, Foster DH: **Role of second- and third-order statistics in the discriminability of natural images.** *J Opt Soc Am* 1997, **14:**2081-2090.
24. Benson PJ: **Morph transformations of the facial image.** *Image Vision Comp* 1994, **12:**691-696.
25. Tolhurst DJ, Troscianko T, Benson PJ, Parraga CA: **The use of morphed images in the quantitative study of human spatial vision.** *J Physiol Lond* 1998, **506:**11-12.
26. Pelli DG, Zhang L: **Accurate control of contrast on microcomputer displays.** *Vision Res* 1991, **31:**1337-1350.
27. Tadmor Y, Tolhurst DJ: **Discrimination of changes in the second-order statistics of natural and synthetic images.** *Vision Res* 1994, **34:**541-554.