The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model

C.A. Párraga [a,*], T. Troscianko [a], D.J. Tolhurst [b]

[a] *Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol, BS8 1TN, UK*
[b] *Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK*

Abstract

Psychophysical thresholds were measured for discriminating small changes in spatial features of naturalistic scenes (morph sequences), for foveal and peripheral vision, and under M-scaling. Sensitivity was greatest for scenes with near natural Fourier amplitude slope, perhaps implying that human vision is optimised for natural scene statistics. A low-level model calculated differences in local contrast between pairs of images within a few spatial frequency channels with bandwidth like neurons in V1. The model was "customised" to each observer's contrast sensitivity function for sinusoidal gratings, and it could replicate the "U-shaped" relationships between discrimination threshold and spectral slope, and many differences between picture sets and observers. A single-channel model and an ideal-observer analysis both failed to capture the U-shape.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Natural scenes; Contrast; Optimisation; Cortex model; M-scaling; Ideal observer

1. Introduction

Classical psychophysical and electrophysiological studies with simple (usually sinusoidal grating) stimuli have resulted in an impressive understanding of channel characteristics of early vision. Yet, it is undoubtedly true that the stimuli viewed in everyday life are different and hugely more complex than gratings. Furthermore, the tasks carried out in everyday life are often more complex than simple grating detection and discrimination tasks. The present paper is motivated by a desire to learn about the relationship between the perception of "natural scenes" and the well-known properties of channels in early vision, as shown with sinusoidal gratings.

It is now generally hypothesised that the organisation of the visual system and the tuning characteristics of individual channels or neurons are optimisations for dealing with the salient information in the natural visual world (Barlow, 1961; Laughlin, 1983; Marr, 1982). The function of red–green colour opponency has been interpreted in these terms (Osorio & Vorobyev, 1996; Párraga, Troscianko, & Tolhurst, 2002; Regan et al., 2001), as has the contrast coding of single neurons or populations of neurons (Clatworthy, Chirimuuta, Lauritzen, & Tolhurst, 2003; Laughlin, 1981; Tadmor & Tolhurst, 2000). The spatial organisation of primary visual cortex (V1) neuron receptive fields seems to match the "statistics" of spatial features in the visual image (Hancock, Baddeley, & Smith, 1992; Olshausen & Field, 1997; Srinivasan, Laughlin, & Dubs, 1982; Van Hateren & Van Der Schaaf, 1998). However, such "visual ecology" generally looks at how the properties of single visual neurons rather than overall visual performance may be matched to the natural

---

environment. We propose that, if the visual system really is optimised for the information in the natural environment, then visual detection and discrimination might be best when we use natural scenes as stimuli or, at least, stimuli with certain statistical characteristics of natural scenes (Geisler, Perry, Super, & Gallogly, 2001; Knill, Field, & Kersten, 1990).

To test this proposal, we need to compare detection or discrimination performance with natural and unnatural stimuli. We use digitised monochrome photographs of natural scenes to represent "natural scene stimuli". But, what is an unnatural stimulus? Natural scenes exhibit many statistical regularities (Geisler et al., 2001; Kersten, 1987), and the Fourier amplitude spectra of natural scenes show a remarkably stable relationship between the spatial frequency (f) and the amplitude of that spatial frequency component (their "second-order" statistics):

$$\text{Amplitude}(f) \propto f^{-\alpha}, \tag{1}$$

where α is the *spectral slope* of the scene and has values close to 1.2 on average (Burton & Moorhead, 1987; Carlson, 1978; Field, 1987; Párraga, Brelstaff, Troscianko, & Moorhead, 1998; Párraga et al., 2002; Tolhurst, Tadmor, & Chao, 1992). Given that natural scenes generally have this property, it is possible to define the degree of naturalness of related stimuli according to how close the α of an image is to its natural, unperturbed value. It is possible to produce versions of images which have α modified by an amount $\Delta\alpha$; such images may be regarded as increasingly "unnatural" as $|\Delta\alpha|$ increases. There have been several psychophysical investigations of visual discriminations using random-dot and digitised photograph stimuli whose amplitude spectra have been manipulated in such a way (Knill et al., 1990; Párraga & Tolhurst, 2000; Tadmor & Tolhurst, 1994; Thomson & Foster, 1997; Tolhurst & Tadmor, 1997).

In this paper, we investigate what is intended to be a more naturalistic discrimination task than has been used before: detection of small spatial changes in stimuli generated by morphing between two natural-scene images (Benson, 1994). Such a task might, for instance, be the basis of identifying facial identity or expressions, or of distinguishing between two slightly different objects. We measure thresholds for discriminating morphed image sequences for stimuli having natural and unnatural second-order statistics, to ask whether human vision *is* optimised for natural statistics. Primarily, we wish to know whether performance in such tasks and the effects of changes in amplitude spectral slope or viewing eccentricity are explicable in terms of the low-level channel structure of the visual system, so well characterised with grating stimuli.

There are a number of image-difference models designed to predict the visibility, e.g., of targets in natural scenes (Daly, 1993; Doll, McWorter, Wasilewski, & Schmieder, 1998; Lubin, 1993; Menendez & Peli, 1995; Rohaly, Ahumada, & Watson, 1997; Watson, 1987; Watson, 2000). The basis of these models is to split the two images to be compared into several spatial-frequency bands (compare Campbell & Robson, 1968; Peli, 1990), weighted by the contrast sensitivity function (CSF) of the observer. For each equivalent pair of points in the images, one must find whether, at each spatial frequency, the difference in contrast between the image patches is at or above the contrast discrimination threshold. This knowledge is provided by knowing the contrast discrimination function (Legge & Foley, 1980). This information needs to be spatially pooled over the whole of the two images, and some process must exist that allows information from different spatial-frequency bands to be combined (Rohaly et al., 1997). Such models have been used to look at applied issues such as image quality (e.g., to evaluate image compression algorithms) or the visibility of small military targets; it is less clear whether they will be able to account for shape discrimination data in experiments such as those proposed here. In this paper, we investigate whether such a multiple frequency-band model can account for the magnitudes of thresholds for the naturalistic morph-discrimination task for stimuli with natural and unnatural second-order statistics, and whether it can account for differences of thresholds between different observers and different viewing eccentricities.

A preliminary account of some aspects of this project has been published (Párraga, Troscianko, & Tolhurst, 2000). Further work has appeared in Abstract form (Párraga, Tolhurst, & Troscianko, 1999; Párraga et al., 2002; Párraga, Troscianko, Tolhurst, & Gilchrist, 2000).

2. Methods

2.1. The natural-image stimuli

The experiments described here are similar in design to those described in our previous work (Párraga et al., 2000). The stimuli were produced from four achromatic images (128×128 pixels, 8 bits of grey level) containing the face of a man, the face of a woman matched for size; and a bull and a car on grey backgrounds. Two different morph sequences were created, one by "morphing" the two faces (called here *man-to-woman*, courtesy of P.J. Benson) into a sequence of 41 slightly different faces, and another by "morphing" the bull into the car (called *bull-to-car*). Both morph sequences then consisted of a series of pictures varying in shape, contrast and texture in small incremental steps of 2.5% steps in the case of the *man-to-woman* series and of 0.5% steps in the case of the *bull-to-car* series. The difference in step size followed preliminary experiments

which showed that the discriminable steps for the *bull-to-car* images were usually less than for the *man-to-woman* images. In both morph sequences, the salient features of the first original image were matched to those of the final original image (e.g., lamps and radiator of the car were matched to the eyes and nose of the bull, etc.). Each image could potentially have represented a real face or object; there were no "ghosts" like those produced in the blending technique of Tolhurst &

Tadmor (Tolhurst & Tadmor, 2000). Fig. 1A shows the original "bull", original "car" and some of the intermediate morphed images. In Fig. 8B, we also report some results collected with two further morph sequences: one of an exaggeratedly smiling face turning in 2.5% steps into an exaggeratedly frowning one, and one of a lemon turning in 2.5% steps into a capsicum/pepper.

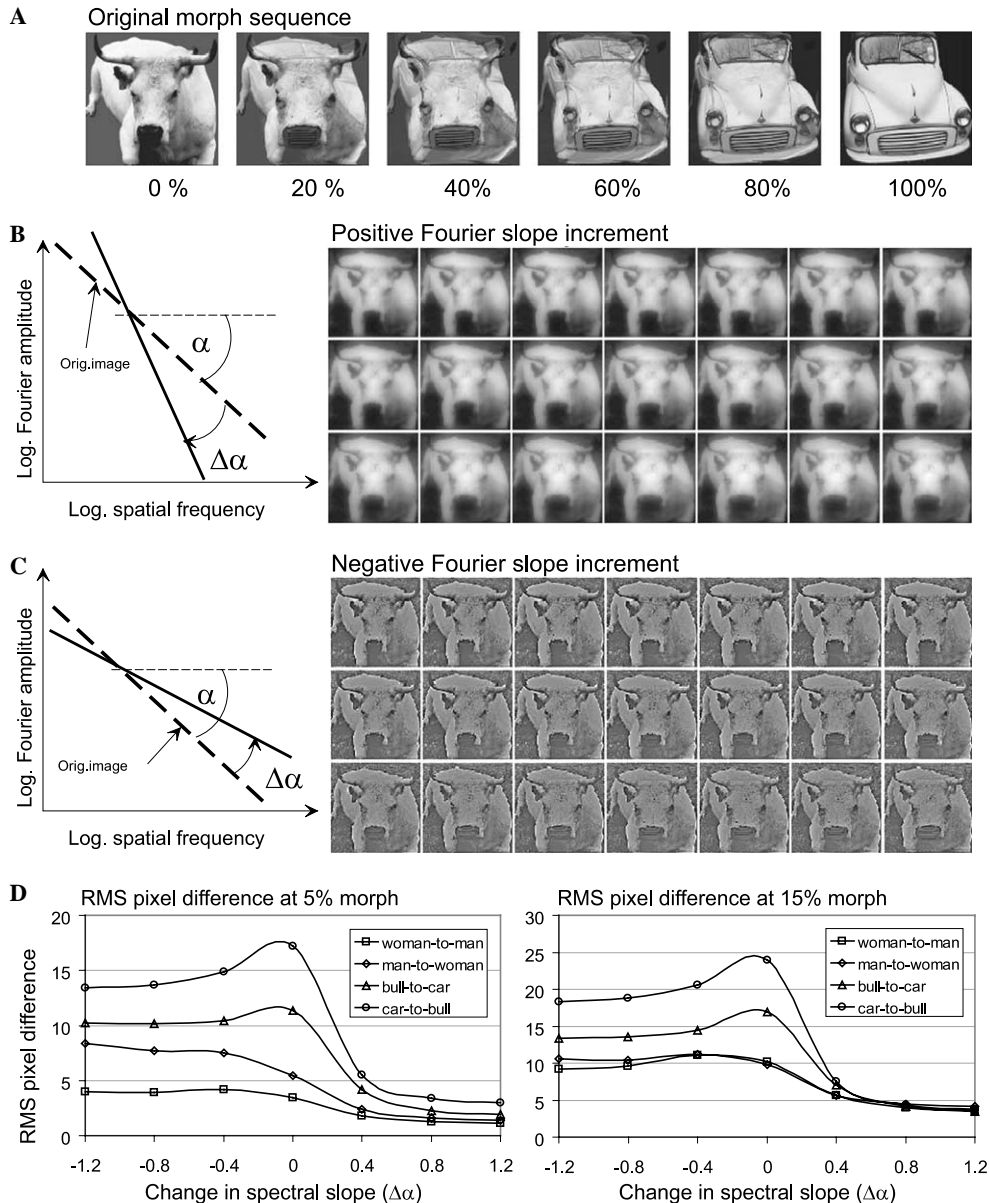These *man-to-woman* and *bull-to-car* image sequences were each used to make seven new image sets which had

Fig. 1. (A) Some examples of the morphed images that were the basis for the stimuli used in this study. A photograph of a bull (left, 0% morph) is gradually morphed into a photograph of a car (right, 100% morph). Some intermediate morphed images are shown: morphed 20%, 40%, 60% and 80% along the scale from "car" to "bull". (B) The graph shows schematically how the amplitude spectrum of an image (dotted line) has its slope (initially α on log/log co-ordinates) increased by an increment Δα to give a new image with steeper slope (solid line). To the right is a sequence of images from the bull-to-car sequence (0–20% morph in 1% steps) after the slopes of all their amplitude spectra has been increased by 0.8. (C) The same as (B) except that a *decrement* of 0.8 in spectral slope is applied (note the reduction in perceived contrast). (D) The root-mean-square pixel difference between the reference image and the 5% morph (left) or the 15% morph (right) for the four morph sequences is plotted against the spectral slope increment of the particular sequence.

the slopes of their Fourier amplitude spectra made steeper or shallower than in the original sequences by small increments or decrements. This was done by multiplying the Fourier amplitude spectrum of each image by a filter of the form

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

where f is spatial frequency and $\Delta\alpha$ takes one of seven values: -1.2, -0.8, -0.4, 0.0, $+0.4$, $+0.8$, and $+1.2$. A positive value is similar to "blurring" of the image (Fig. 1B); a negative value represents image "whitening" or edge enhancement (Fig. 1C), which is usually accompanied by a fall in overall contrast (Párraga et al., 2000; Tadmor & Tolhurst, 1994). An increment of 0.0, of course, represents the original image sequence.

After filtering the spectra of the images, the filtered images were obtained by inverse Fourier transformation. The original images had grey levels in the range 0–255, but the filtering changed that range, especially for the negative slope increments, where the range became greater and included negative numbers. Since only positive light levels are possible and since the VSG display (see below) is capable of displaying only 256 grey levels at a time, it was necessary to scale the pixel values of the filtered images. The images in a single set of 41 (one pair of original photographs, and one spectral slope change) were scaled as one unit. First, a constant was added to or subtracted from all floating-point pixel values to make the smallest one in the set equal to zero. Then, the pixel values of each image were stretched or compressed so that 256 integer grey levels encompassed the whole set, and the darkest pixel in the whole set was 0 while the brightest was 255. Different scaling factors were applied for the seven spectral slope sets for each original image pair.

On average, the slope (α in Eq. (1)) of the amplitude spectrum of a natural image is about 1.2 (Tolhurst et al., 1992) but the four images used here had steeper slopes (approximately 1.3–1.5), which is our experience of "portrait-like" images of single objects on a blank background (in fact these sequences are the same as those used in Párraga et al., 2000; see Fig. 2 of that paper to see a plot of its Fourier amplitude). The amplitude spectra of the originals were not necessarily exactly straight and the slopes of the images in a morphed set were not necessarily the same. We did not normalise within or between image sets in order, for instance, to ensure that pictures all had the same power. The left side of panel D in Fig. 1 shows the root-mean-square difference (ΔRMS) between the pixel values of first picture (reference) and the picture corresponding to 5% morph change (test) for each of the four morph sequences and the seven values of $\Delta\alpha$ used in this study. The right of panel D shows the same for the picture corresponding to 15% morph change. Here ΔRMS contrast difference is defined as

$$\Delta\text{RMS} = \sqrt{\sum_i \frac{(R_i - T_i)^2}{n}}, \tag{3}$$

where R_i represents the ith pixel of the reference picture and T_i represents the same for the "test" picture of the same morph sequence. We have shown elsewhere that the differences in apparent contrast between stimulus sets with different slope increment (evident in Fig. 1) do not greatly influence the forms of the psychophysical results (Párraga et al., 2000; Tolhurst & Tadmor, 2000). Previously (Párraga & Tolhurst, 2000; Tadmor & Tolhurst, 1994; Tolhurst & Tadmor, 2000), we *have* applied constraints to image sets, but with the result that none of the images ever has a strictly natural appearance.

2.2. Experimental conditions

Pictures were presented in the centre of a Sony Trinitron monitor screen driven by a Cambridge Research Systems VSG 2/4 Graphics Card. This had pseudo-15-bit control (Pelli & Zhang, 1991) of the luminance of the pixels so that it was possible to compensate for first-order luminance nonlinearities in the display and still present stimuli (including very low contrasts sinusoidal gratings) to a full precision of 256 grey levels. We did not attempt to account for the effects of pixel neighbours along the raster lines (Garcia-Perez & Peli, 2001; Klein, Hu, & Carney, 1996; Pelli, 1997; Pelli & Zhang, 1991; Schofield & Georgeson, 1999; Schofield & Georgeson, 2000). The screen measured 36.0×29.5 cm and was viewed from 2 m, so that it subtended $10.26 \times 8.41°$. The 128 pixel square images usually measured 8.5 cm ($2.43°$) square, and each pixel measured 1.14 min; each "logical" image pixel occupied a 2×2 square of hardware display pixels. To avoid spurious cues resulting from edge effects, all pictures were smoothed at the edges with a Gaussian roll-off ($SD = 15$ pixels). To make smaller versions of the images ("*small foveal*" experiments) the 128×128 pixel full-sized images (including the roll-off) were subsampled by the VSG "moverect" command, with the memory size of source and destination being different. The small pictures consisted of 90 times 90 pixels, each measuring 0.57 min (the hardware resolution).

The screen had a mean luminance of 85 cd/m² in all parts not occupied by the stimuli. The brightest pixel in an image never exceeded a luminance of 170 cd/m² (double the screen background) while the darkest might nominally have been 0 cd/m². The mean luminances of the stimulus images were not necessarily 85 cd/m². The frame rate was 80 Hz.

Observers viewed monocularly and freely for foveal viewing (Párraga et al., 2000; used binocular viewing) but, for peripheral viewing, they fixated monocularly upon a red light-emitting diode (LED) at 3 or 6° along the horizontal axis towards the side of the screen. In
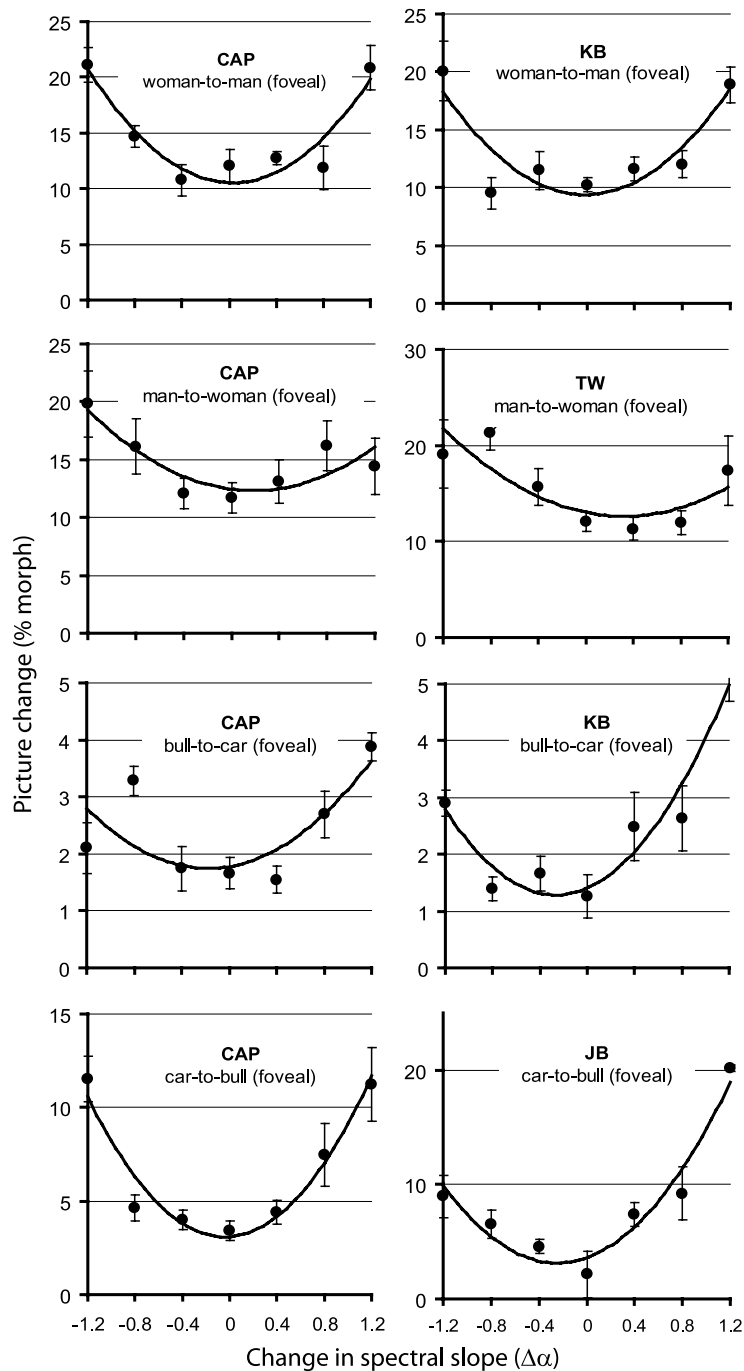
Fig. 2. Discrimination thresholds for monocular foveal viewing of the four main sets of morphed images. Threshold is plotted as percentage morph needed for discrimination against the change of amplitude spectral slope. ±1 standard error is shown. The solid curves are the best-fitting second-order polynomials, fitted by minimising χ^2 (i.e., the residual sum of squares weighted by the standard errors of the experimental measurements), which is less than 5 for all fits, except 13.08 for CAP bull-to-car and 6.66 for KB bull-to-car; adding a third-order term cause a reduction in χ^2 by less than 3.0 in all cases, except 3.18 for CAP bull-to-car and 3.17 TW man-to-woman. Results for observer CAP on all 4 morph sets are shown on the left; for several other observers on the right.

peripheral viewing, images were presented in the nasal hemifield. The better eye was chosen in all cases (the preferred eye for shooting or for looking through a telescope).

There were four different experimental conditions: in condition 1, viewing was foveal and the images mea-

sured 2.43° square at the eye. In conditions 2 and 3, viewing was with the nasal hemifield, with the centres of the 2.43° square images at 3° and 6° into the nasal hemifield, respectively. In the fourth condition, viewing was again foveal, but the images were reduced in size by a factor of 0.37 to be 0.9° square at the eye. This

reduction in image size was intended to "M-scale" (Drasdo, 1991; Levi, Klein, & Aistebaomo, 1985; Rovamo & Virsu, 1979) the foveal images relative to the 6° peripheral ones. It is argued that stimuli can be scaled in size to compensate for putative differences in visual acuity and/or in cortical magnification between fovea and periphery. There are many different ways to calculate an M-scaling factor, depending upon assumptions and, perhaps, depending upon the kind of task (Levi et al., 1985; Tolhurst & Ling, 1988). We used a conservative estimate of 1:2.7 here (Rovamo & Virsu, 1979) whereas Tolhurst and Ling (1988) argued that actual cortical magnification rather than acuity would change 7.4 times rather than 2.7 times as we move from fovea to 6° peripherally. We shrank the foveal picture by a factor of 2.7 rather than magnifying the peripheral one by the same factor, since parts of a greatly enlarged image centred at 6° would be less than 3° from the fovea, and discriminations might then be performed with less eccentric parts of the visual field.

2.2.1. Observers

All experiments were carried out on two main observers: CAP (one of the authors) was a well corrected myope, while KB (a naïve but experienced observer) had normal vision. The detailed modelling of the discrimination thresholds was carried out for these two observers. Most of the experimental observations were confirmed on up to 4 other naïve observers, who each completed many of the experiments; they were students at Bristol University, and all scored normally on a Snellen acuity test at the same viewing distance as used in the experiments. We also include for new analysis the results for 2 other observers who contributed to our previous report (Párraga et al., 2000), and Fig. 8B shows data for the "good eye" of each of 6 amblyopic observers reported elsewhere (Tolhurst & Párraga, 2003).

2.3. Experimental protocols

The observers had to discriminate between the original (*reference*) and morphed (*test*) images belonging to the same morph sequence and with the same slope-increment in their amplitude spectra (and thus sharing similar second-order statistics) in a modified 2AFC procedure, using a conventional staircase technique. In a single trial, the observer was presented with *three* images sequentially (each presented for 500 ms with intervals of 200 ms between them). The second presentation was always known to be a copy of the reference image; this reference interval is needed in complex visual discriminations otherwise observers require very detailed memory of the various stimuli. The computer chose randomly whether to present a second copy of the reference for the first time interval and the test in the third time interval, or vice-versa. The observer had to decide whether

the "odd one out" (morphed test image) was in the first or the third presentation in the trial, and indicated their choice to the computer by pressing either the left or the right mouse button. Auditory feedback was given as to whether the choice was correct. The same test and reference images were presented five times. If the response was correct all five times, the task was made harder (by selecting a test picture closer to the reference in the morph sequence). If the observer made one or more errors in the sequence of five trials, then an easier morph image was chosen for the subsequent five trials. The upward and downward steps in the staircase were the same size, and stepsize remained constant throughout the procedure. In fact, two independent staircases were interleaved randomly for each morph sequence, one starting from the bottom (difficult task) and becoming increasingly easier and another starting from the top (easy task) and becoming more difficult. After the staircases had stabilised, psychometric functions were fitted to the data pooled from the two staircases (typically 100 trials per staircase).

In a typical experiment, four different morph sequences (with different spectral slopes, but from the same pair of original photographs) were randomly interleaved, and the observer's thresholds for these were measured concurrently. For the *man-to-woman* sequence, we performed two sets of experiments: one with the "man" as the reference, and one with the "woman" as the reference (called *woman-to-man*). Similarly, we used both "bull" and the "car" as the reference in different experiments (*car-to-bull* as well as *bull-to-car*).

Each psychometric function was fitted with the integral of a normal distribution, which was constrained to lie between 50% (guess rate in a 2AFC) and 98% (allowing for a 2% "finger error"—the chance that an observer might sometimes push the wrong response button by mistake). The discrimination 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 the trials. The slope and position (threshold parameter) of this cumulative normal were estimated using a SIMPLEX routine, which maximised log-likelihood (Press, Flannery, Tekulosky, & Vetterling, 1986). Standard errors for the discrimination thresholds were estimated from the inverse of the second differential of the likelihood function at the maximum of the merit function (Edwards, 1972). We confirmed, by computer simulation of staircases, that these estimated standard errors did describe the range of estimated thresholds returned by multiple staircases.

2.4. Sensitivity to sinusoidal gratings

The observers' CSF were measured under analogous conditions: with vertical sinusoidal gratings in a square window of 2.43° for foveal, 3° eccentric and 6° eccentric

viewing, and in a small square window of 0.9° for "small-foveal" viewing. Experiments were again performed monocularly, and there was a Gaussian roll-off around the four sides of the grating square. Thresholds were estimated using a 2AFC conventional staircase: the observer had to indicate in which of two 500 ms intervals a grating was presented. The contrasts of gratings of different spatial frequencies were increased or decreased every five trials on the basis of how many correct responses were made (as above). A cumulative normal curve was fitted to the psychometric function to estimate contrast threshold by interpolation.

2.5. The contrast at a point in a natural image

We present a model to explain the magnitudes of the thresholds for discriminating between morphed natural scenes. The model is based on knowledge of primary visual cortex and has much similarity with others (Daly, 1993; Doll et al., 1998; Rohaly et al., 1997; Watson, 1987; Watson, 2000). These models suppose that a visual image is initially processed in parallel by channels or neurons with different optimal spatial frequencies but all with much the same bandwidth of about 1 octave (Blakemore & Campbell, 1969; De Valois, Albrecht, & Thorell, 1982; Movshon, Thompson, & Tolhurst, 1978; Tolhurst & Thompson, 1981; Watson & Robson, 1981). Thus, as a precursor to modelling how the visual system compares two slightly different images, we first calculate the contrast at each point in an image, at each of several spatial frequency scales (Párraga & Tolhurst, 2000; Peli, 1990; Tadmor & Tolhurst, 1994; Tolhurst & Tadmor, 1997). We define contrast at the point $[x, y]$ and in the frequency band F as

$$C_F(x, y) = \frac{a_F(x, y)}{l_F(x, y)}, \qquad (4)$$

where $a_F(x, y)$ is a bandpass filtered version of the original image convolved with a circularly-symmetric filter with frequency response given by

$$A_F(f) = \exp\left[-\frac{(f - F)^2}{2\sigma^2}\right], \qquad (5)$$

where f is spatial frequency, while σ is the spread of the Gaussian frequency–response curves and is chosen to be $0.3F$ so that the bandpass filters have a bandwidth of about 1 octave. $l_F(x, y)$ is the result of convolving the original image with a circularly-symmetric low-pass operator with frequency response given by

$$L_F(f) = \exp\left[-\frac{(f)^2}{2\sigma^2}\right]. \qquad (6)$$

Division of a_F (the bandpassed convolution) by l_F (the local mean luminance) is a model of the fact that the visual system encodes contrast rather than lumi-

nance per se (Shapley & Enroth-Cugell, 1984); the mean luminance is calculated over an area proportional to the period of F. Others (Brady & Field, 2000; Field, 1994; Van Hateren & Van Der Schaaf, 1998) model contrast encoding by taking the logarithm of the pixel values before applying linear filtering operations to images. Usually, we have calculated contrast at each point in an image within five spatial frequency bands, one octave apart.

To model how the visual system compares two images, we calculate $C_F(x, y)$ for both images at all frequency scales, and then we compare the contrasts in the two images, *point by point* within each frequency band (see Section 3 for details). In previous papers, we averaged the contrast across the image within each frequency band before comparing that single value with the single averaged-contrast value of another image. That may have been appropriate when the experimental variable changed the power or contrast over the *whole* image (Párraga & Tolhurst, 2000; Tadmor & Tolhurst, 1994; Tolhurst & Tadmor, 1997), but is inappropriate here where the differences between stimuli involve changes in the shape or contrast of *spatially localised* features.

As well as the morphed images of natural scenes, we have modelled "images" of sinusoidal gratings of known Michelson contrast to be able to express contrast at each point in the morphed image as *equivalent Michelson contrast*: the contrast of optimal grating that would evoke the same "response" as that location of the image. This allows us to relate the contrasts in images to measurements of an observer's contrast thresholds for detecting gratings and for discriminating differences in contrast between pairs of gratings.

3. Results

3.1. Psychophysical results: Foveal viewing

Fig. 2 shows the discrimination thresholds for eight different experiments with monocular foveal viewing (condition 1—see Section 2). The discrimination thresholds are expressed as the percentage morph that is just discriminable in a 2AFC, and are plotted along with ± one standard error against the change in amplitude spectral slope. A spectral change ($\Delta\alpha$) of zero corresponds to unblurred and unwhitened (i.e., natural) scenes. The left-hand panels of the Figure show the results for observer CAP on the four different morph sequences, and the right-hand panels show experiments on the same 4 morph sequences but on different observers. The photographs of the man and the woman were well matched so that successive 2.5% morph steps were difficult to discriminate; discrimination thresholds for all observers are in the range 10–20% for a $\Delta\alpha$ of zero. The bull and car photographs (Fig. 1A) were more different, so that a

2.5% morph step was more easily discriminable, and thresholds were actually measured with morph sequences that differed successively by only 0.5%. The discrimination thresholds are low: in the range 1.5–5% with the thresholds at the bull end of the sequence (third row of Fig. 2) being lower than those at the car end (bottom row of Fig. 2).

Discrimination thresholds are low in the mid range flanking spectral slope changes near zero ("natural" scene statistics), and are highest at extreme negative values of $\Delta\alpha$ (image "whitening") or at extreme positive values (image "blurring"). In seven out of the eight examples, the lowest threshold is at a $\Delta\alpha$ of -0.4, zero or $+0.4$. This confirms our preliminary report (Párraga et al., 2000) and is similar to results with spectrally blended image pairs (Tolhurst & Tadmor, 2000). The data, which follow a roughly "U-shaped" course, could often be described by second-order polynomials; the solid lines in Fig. 2 shows the best-fitting second-order polynomials, fitted by minimising χ^2 (Press et al., 1986). In the examples of Fig. 2, the minima of the U-shaped polynomials generally correspond to the region where the amplitude spectral slope is unmodified (natural scenes).

In all, we performed 26 such experiments with monocular foveal viewing, involving up to six observers on some or all of the four morph sequences. Additionally, two further observers (other than CAP) performed five binocular experiments between them for our previous study (Párraga et al., 2000). We attempted to learn how many of the experimental data sets could be described by some generic "U"-shape and where the minimum of that "U" might be; we fitted the 31 data sets with second-order polynomials, and compared the goodness of fit with lower (straight line) and higher-order polynomials. The second-order polynomial fit has 3 degrees of freedom, and 13 of the experiments had χ^2 less than 9.49 (the critical value for $P = 0.05$) with the median value being 10.47. The inclusion of the second-order term compared to a 2-parameter straight line fit caused a large reduction in χ^2 (and a loss of 1 degree of freedom) in all but three of the 31 experiments; the median fall in χ^2 was 26.38. On the other hand, addition of a fourth parameter (a third-order term) caused no significant improvement to the fit in 23 cases; the median fall in χ^2 caused by adding the third-order term was only 1.49. The comparison of first-, second- and third-order polynomial fits shows that the experimental data mostly fall on a function with a single minimum in the mid $\Delta\alpha$ range. An example where a second-order polynomial was not a good fit at $P = 0.05$ and where addition of a third-order term caused a near-significant improvement is shown in Fig. 2 (CAP *bull-to-car*, see legend for χ^2 values). Overall, although a second-order polynomial might not have been an ideal fit, 27 out of 31 experiments showed lowest thresholds for $\Delta\alpha$ values in the

mid range. One of the four experiments that failed to conform can be seen in Párraga et al. (Párraga et al., 2000: Fig. 3, open triangles, bottom left panel).

The minimum of the best fitting second-order polynomial is an indication of the second-order image statistics at which observers are best able to discriminate the spatial structure of the images. The distribution of these minima for foveal viewing is shown as the open bars in Fig. 5A. In almost all the experiments, discriminations were most acute for $\Delta\alpha$ values within ± 0.4 of "natural statistics". The polynomials had minima with a median at a spectral slope increment of $+0.096$ and a mean of $+0.049$ (*SE* 0.061), not significantly different from zero.

Given that the y axis units in Fig. 2 (% morph) do not represent any universal (physical) measure of change in different sequences (for example, 10% change in the woman-to-man sequence might be equivalent to 2% change in the bull-to-car sequence) it is illustrative to re-plot some of the data in terms of ΔRMS as defined in Section 2 (Eq. (3)). In particular, Fig. 1D shows that ΔRMS changes most with percentage morph at $\Delta\alpha = 0$, when the thresholds expressed as percentage morph are lowest. This raises the question whether the different thresholds at different $\Delta\alpha$ values result simply from differences in ΔRMS. Fig. 3A shows the ΔRMS corresponding to the threshold values for four sets of experimental results (shown on the left side of Fig. 2). The other panels in Fig. 3 replot three of the experimental sets (filled circles) while the open circles show the discrimination thresholds that would be predicted if the observer were discriminating purely on the basis of RMS pixel change in each of the morph sequences. These model values were obtained by calculating the actual ΔRMS present in the just-discriminable image pair for $\Delta\alpha = 0$, and then computing the percentage morph change that produced the same value of ΔRMS in all the other sequences.

The plots in Fig. 3 are representative of the rest of the dataset and show that ΔRMS alone cannot explain our psychophysical results. Although the ΔRMS metric makes the thresholds for the *bull-to-car* sequences similar in magnitude to the *man-to-woman* and *woman-to-man* sequences, the ΔRMS values for the *car-to-bull* are rather different (Fig. 3A). In some cases, especially for sequences that were "whitened" ($\Delta\alpha = -0.4$ and $\Delta\alpha = -0.8$ in panels b and d) it seems that changes in RMS pixel value do provide a clue for the observers to discriminate changes in the morphs. However, this is not true for any of the "blurred" sequences, where ΔRMS underperforms seriously.

3.2. Ideal observer analysis

Figs. 3B–D also show the thresholds predicted by an ideal observer analysis (open triangles). The estimate of the ideal observer performance was calculated in two
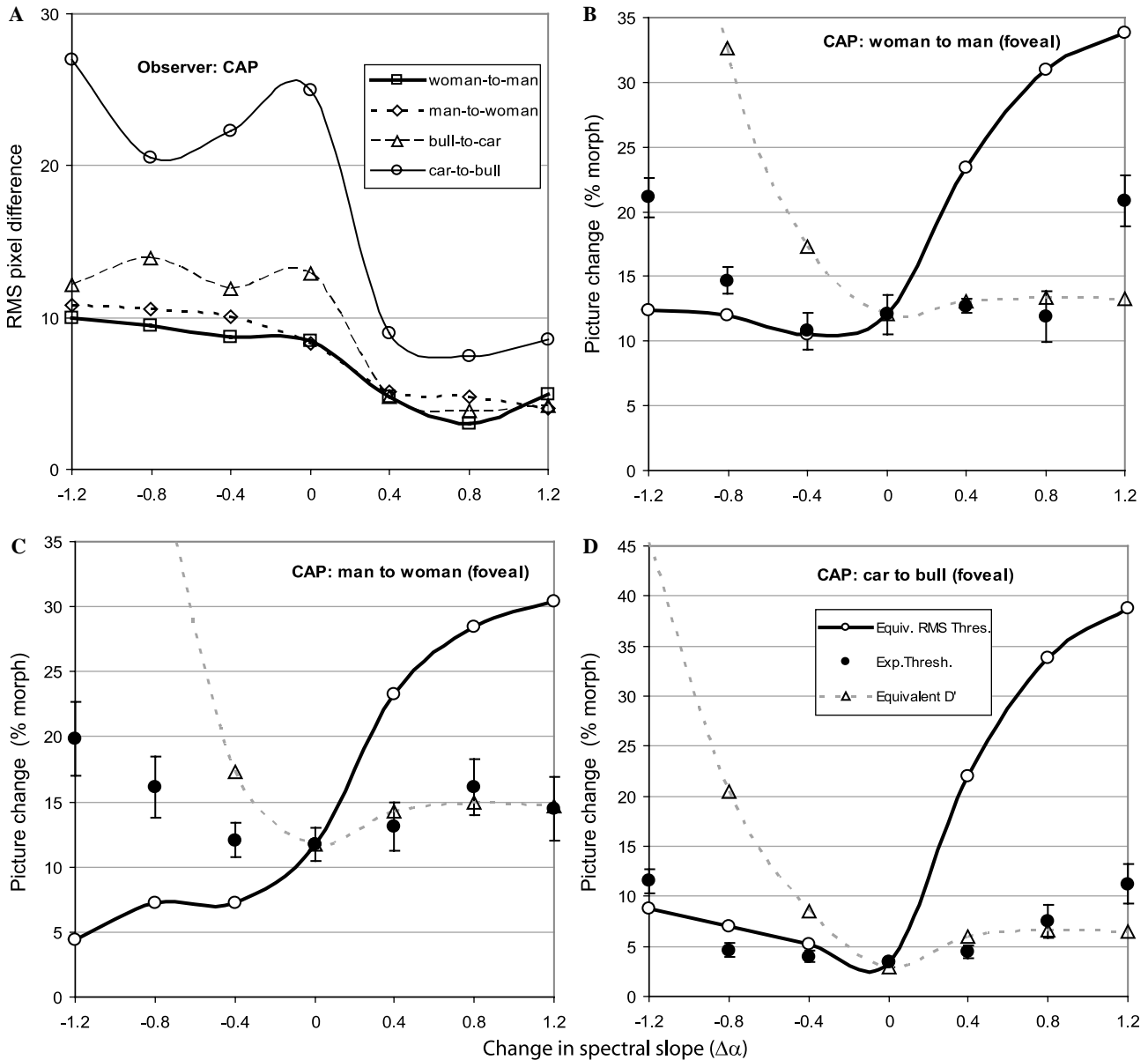
Fig. 3. (A) The four experimental datasets from the left of Fig. 2 are replotted to show threshold measured as RMS pixel difference. (B–D) Three of those datasets are plotted separately as filled circles. The solid lines and open circles show predictions of the thresholds if the observer had been detecting a fixed change in RMS pixel value. The dotted lines and open triangles show an ideal observer prediction of the thresholds.

stages. First, observer CAP's CSF was measured for sinusoidal gratings, and a corresponding "eye point-spread function" calculated (as the inverse Fourier transform). Each picture in the morph sequences was then convolved with this point circularly-symmetric spread function and the value of d' (Geisler, 2003, his Eq. (16)) was calculated for every pair of reference-test pictures using the equation

$$d' = \sqrt{2 \sum_i \frac{(r_i - t_i)^2}{(r_i + t_i)}}, \qquad (7)$$

where r_i and t_i represent the ith pixel of the convolved reference and test images. It has also been suggested to

us that it would be more appropriate to square the denominator of Eq. (7) (in agreement with a light-adaptation observer); however, this made very little difference to the forms of the plots in Fig. 3.

The criterion for discrimination in any particular experiment was defined as the value of d' corresponding to the just discriminable pair of pictures in the "natural" condition ($\Delta\alpha = 0$). The triangles in Figs. 3B–D correspond to the percentage morph change necessary so that d' reaches the criterion in all $\Delta\alpha$ conditions. The ideal observer analysis predictions badly underperform on the "whitened" ($\Delta\alpha < 0$) side but do better on the "blurry" ($\Delta\alpha > 0$) side. This is the opposite of what occurred with RMS pixel difference threshold, and can be

explained in terms of the nearly complete removal of useful information produced by both whitening the pictures and convolving them with the point-spread function (blurring). In summary, Fig. 3 shows that the experimental results obtained for one observer (CAP) cannot be explained in simple terms by either detection of RMS pixel changes or signal-to-noise measures (such as d') and a more complex model is needed. The same applies to the other morph sequences and observers.

3.3. Psychophysical results: Peripheral viewing

Fig. 4 shows four examples of the discrimination threshold data for two of the morph sequences obtained for two subjects (KB and CAP) with monocular peripheral viewing: with images centred 3° eccentric (open circles) and 6° eccentric (filled circles). For comparison, the
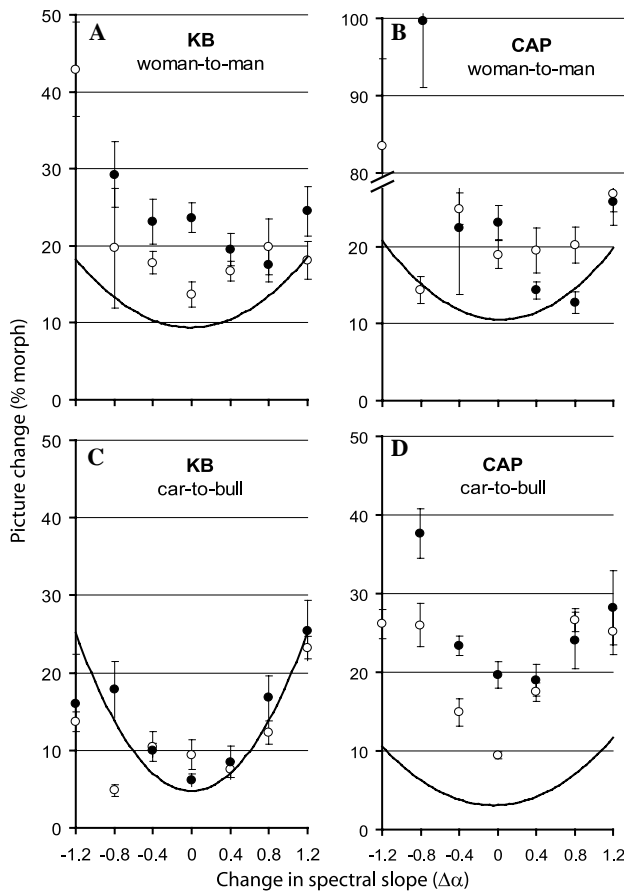

Fig. 4. Four examples of the effects of monocular peripheral viewing on the magnitudes and forms of the results. The open symbols show the discrimination thresholds (±1 standard error) when the observer fixated 3° from the centre of the 2.43° square images; the filled symbols are for 6° eccentric viewing. The solid curves are the best fitting second-order polynomials fitted to the equivalent *foveal* results; three of these lines can be found in Fig. 2. Data for two observers and two morph sequences are shown. Note the break in the ordinate in the top right panel, and note that some data are missing at Δα of −1.2 because the thresholds were too high to measure.

solid curves are the second-order polynomials fitted to the equivalent *foveal* data (three of the lines can be found in Fig. 2). The two upper graphs show the most usual behaviour: a general increase in the discrimination thresholds as the stimuli are presented more peripherally. The bottom graphs are less typical, either because discrimination thresholds do not seem to change when the stimuli are presented peripherally (bottom left) or the threshold increment is larger than general (bottom right). Very often, the thresholds for a Δα value of −1.2 (extreme image "whitening) were very high (e.g., open circles, top graphs) and, indeed, some of the data points for 6° eccentric viewing are not shown in the graphs (Δα = −1.2 filled circles) because the thresholds were so high that they could not be measured. In summary, discrimination thresholds for peripheral morph stimuli are generally elevated, especially at extremely negative values of Δα (image whitening). However, the data do seem to continue with the trends shown in Fig. 2, although some curves are less obviously "U-shaped".

We fitted second-order polynomials to the peripheral data. Analysis of the positions of the minima of the "U-shaped" curves reveals little difference when the stimuli are shown foveally and at 3° eccentricity. The solid black bars in Fig. 5 shows the distributions of the minima (Fig. 5A) and the mean and standard error of the minima (Fig. 5B) for 3° viewing, for comparison with foveal viewing (open bars). Sometimes the polynomials were fit only to the 6 data in the range of Δα from −0.8 to +1.2 since a very high threshold at Δα of −1.2 would distort the fit, causing the minimum to shift towards higher Δα values. The grey bars in Fig. 5 show the distribution and summary of the minima for 6° eccentric viewing. It was often necessary to ignore high or absent thresholds at Δα of −1.2 or even −0.8 to get a reasonable fit of a second-order polynomial to the remaining data; these high values distorted the polynomials so that their minima occurred at Δα values that were obviously too positive. Even after removing the high threshold values, the minima for 6° viewing are still significantly shifted towards positive values (see the filled symbols in three of the four panels of Fig. 4). The mean of the minima is at a positive slope increment of 0.25 (significantly different from zero; $SE = 0.10$; $t = 2.5$; $n = 16$; $P < 0.05$).

3.3.1. M-scaling of the stimuli

For two observers (KB and CAP), we examined the effects of putative M-scaling on the thresholds for peripheral viewing. An elevation of threshold for peripheral viewing might perhaps be compensated by making the peripheral images larger. In fact we *reduced* the size of the foveal images by a factor of 2.7 (Rovamo & Virsu, 1979) to accomplish M-scaling (see Section 2). Fig. 6A shows the measured contrast sensitivity functions for
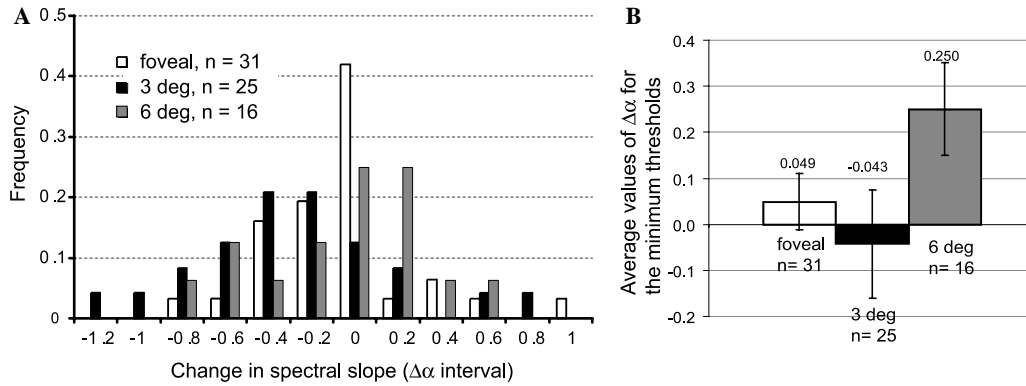
Fig. 5. (A) A summary of the results for foveal and peripheral viewing of 2.43° square images, showing the distributions of the spectral slope where a fitted second-order polynomial was at a minimum. Foveal—open symbols; 3° peripheral—black symbols; 6° peripheral—grey symbols. (B) The means of the three distributions (for foveal and two peripheral eccentricities) are shown ±1 SE. The means for foveal and 3° viewing are within 1 SE of zero. The mean for 6° viewing is significantly different from zero ($t = 2.5$; $P < 0.05$).
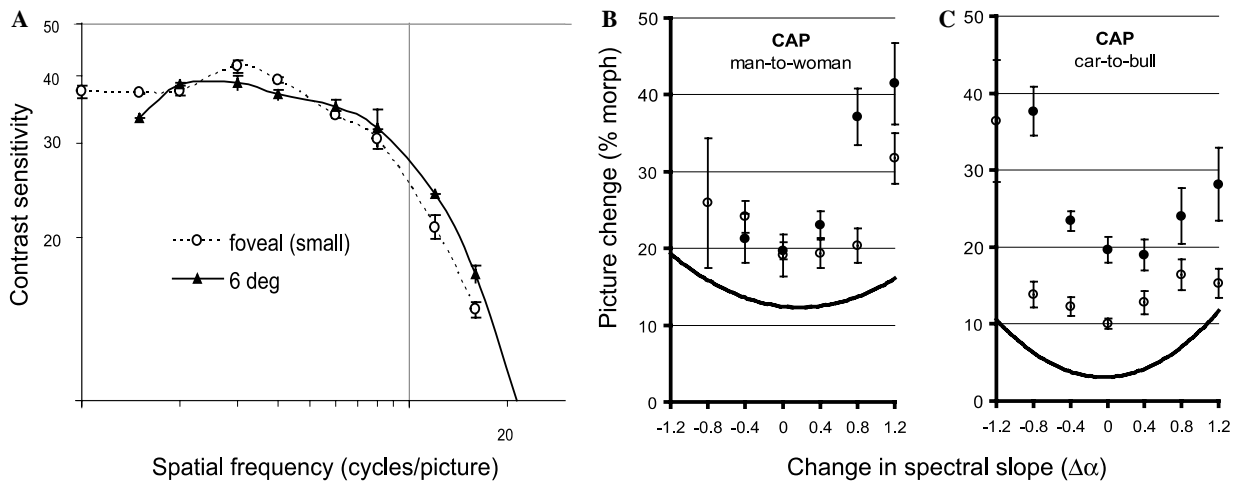

Fig. 6. (A) The effects of M-scaling on the contrast sensitivity to sinusoidal gratings: open circles and connecting dashed lines measured with foveal viewing of square patches of gratings measuring 0.9° square; filled symbols and connecting solid lines measured with 6° eccentric viewing of gratings 2.43° square. Note that the CSFs overlap when spatial frequency is expressed in cycles per picture. (B) and (C) Two examples of the effects of M-scaling on thresholds for discriminating morphed images. Open symbols are for small foveal images, filled symbols are for normal sized 6° peripheral ones; solid line is the second-order polynomial fitted to the foveal thresholds for normal-sized images.

sinusoidal gratings for observer CAP using both M-scaled (0.9° square images) gratings viewed foveally and 2.43° square gratings viewed peripherally. Contrast sensitivity (the inverse of the lowest Michelson contrast needed for the observer to detect a sinusoidal grating) is plotted against the spatial frequency *expressed as cycles/picture*, and not as cycles/degree. The shape and position of the two CSFs are very close showing that the M-scaling factor can indeed compensate for the observer's differences in *grating* acuity.

Figs. 6B and C show two examples of the effects of M-scaling on morph discrimination thresholds for the same observer. The filled symbols are the thresholds for morph discriminations with 6° eccentric viewing of 2.43° square images (the filled symbols in Fig. 6C can be found in Fig. 4); the solid lines are the second-order polynomials

fitted through the results for foveal viewing of the same 2.43° square images, showing that peripheral viewing generally raises thresholds. The open circles in Figs. 6B and C show the foveal thresholds for discriminating changes in M-scaled 0.9° square images. Reducing the foveal images in size by a factor of 2.7 has indeed elevated the discrimination thresholds, as expected. As is also common with the 6° data, the foveal thresholds for $\Delta\alpha$ of -1.2 sometimes became unmeasurable when the images were small. M-scaling has compensated over part of the $\Delta\alpha$ range in Fig. 6B, but not at the most negative values of $\Delta\alpha$. In the experiment in Fig. 6C, M-scaling has not compensated nearly enough. Overall, M-scaling has moved the foveal results towards the 6° peripheral ones, but generally has not compensated nearly enough. The discrimination thresholds produced for the size-reduced

stimuli were still fitted by the same kind of "U-shaped" template as the rest of the results, but with the minima positioned to the right side of the plot (mean value of $\Delta\alpha = 0.266$, $SE = 0.045$, $n = 9$), sharing some characteristics with the 6° eccentricity data. The minima are at an α value significantly different from zero ($t = 5.91$; $P \ll 0.01$).

3.4. A spatial frequency channel model for natural-scene discrimination

We have applied a multiple-channel model (Rohaly et al., 1997; Watson, 1987) to estimate the visibility of differences between the test and reference pictures for all the experiments performed by the two main observers (KB and CAP) in all viewing conditions, as well as a more limited analysis of the results of the other observers. The band-limited contrast (Peli, 1990) was calculated at each $[x, y]$ location in the test and reference images at each of several one-octave spatial frequency bands, F, to give $C_F(x, y)$ (see Section 2, Eq. (4)). Subsequent calculations were performed on the central 68×68 pixels (out of 128×128) of the contrast-arrays. If two pictures are slightly different, then we would expect that the contrast would be slightly different in the two pictures at one or more locations, and in one or more spatial-frequency bands. The model must determine whether these differences in contrast are sufficient to allow the observer to discriminate between the pictures.

The first stage is to estimate how each contrast difference at each location and in each frequency band might contribute *individually* to discrimination. We calculate the absolute value of the difference in contrast between the two pictures under comparison at each location and in each frequency band

$$\Delta C_{F,j}(x, y) = |C_{F,j}(x, y) - C_{F,0}(x, y)|, \qquad (8)$$

where j is the picture number (1–40) of the test stimulus and $j = 0$ represents the reference picture. Then, we estimate how much each value of ΔC might contribute towards the visibility of the difference between the pictures, by evaluating each ΔC value against the familiar "dipper function" for contrast discrimination for sinusoidal gratings (Legge, 1981; Legge & Foley, 1980; Nachmias & Sansbury, 1974). Each value of $\Delta C_F(x,y)$ is treated as if it is the contrast increment of a sinusoidal grating of frequency F to be compared against a reference or pedestal grating whose Michelson contrast, $\bar{C}_{F,j}(x, y)$, is the average of the paired contrast values in the two pictures at that location and frequency band.

$$\bar{C}_{F,j}(x, y) = 0.5|C_{F,j}(x, y) + C_{F,0}(x, y)|. \qquad (9)$$

The observers' contrast *discrimination* functions for gratings were estimated indirectly by adjusting the position on the x-axis (contrast reference) and y-axis (contrast difference) of a "dipper function" template

(Fig. 7A) for contrast discrimination according to the observers' contrast *detection* thresholds measured for a similar grating (Párraga & Tolhurst, 2000). Thus, the model dipper functions were determined from each observer's CSFs at each of the viewing eccentricities and picture sizes. Any differences between observers'
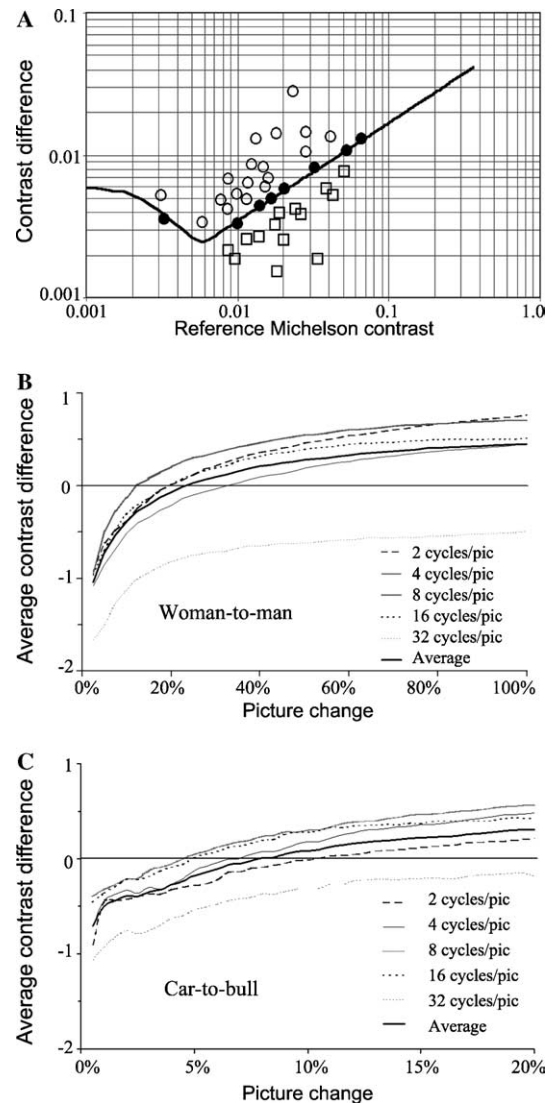

Fig. 7. (A) A schematic of the "dipper function" template used in the model; the template is moved on the x and y axes so that the y-axis intercept and the x-axis value of the dip are the same as the observer's *detection* threshold for gratings of that spatial frequency. The open circles show schematic data for pictures that should be easily discriminated; the squares show data for pictures that should not be discriminable, while the filled circles show data for a pair of pictures that might be just at threshold. (B) The average of the logs of the visibility values V in each frequency band is plotted against the percentage morph change between the reference image (0%) and each test image. Results for observer CAP viewing foveally the $\Delta\alpha = 0$ set of *woman-to-man* images. The horizontal line at an average cue of zero indicates one hypothesis as to when the contrast-difference cues might become visible. (C) The same as (B) except for the *car-to-bull* images. Note that *two* frequency bands cross the zero-criterion line together.

abilities to discriminate between pictures or any effects of different eccentricities should hopefully be accounted for by differences in their CSFs.

Fig. 7A shows the general shape of the dipper template. Note that the linear, "Weber" part of the function has a slope of only 0.7 on log/log axes in our experiments (Legge, 1981). Given a value of $\bar{C}_F(x, y)$, we determine from the dipper function for frequency F the theoretical just-noticeable difference in contrast for real gratings, $D_F(\bar{C}_F(x, y))$. A measure of the visibility of the contrast difference in the two pictures at that location is then given by

$$V_F(x, y) = \Delta C_F(x, y)/D_F(\bar{C}_F(x, y)). \tag{10}$$

When expressed as a logarithm, this is the distance of a calculated contrast value above or below the dipper template.

3.5. Pooling rules

The second stage in the model is to pool the many cues, V, provided at different locations and different frequency bands to give an overall assessment of whether or not the two pictures differ sufficiently for discrimination to be made. In general, we might expect that, if the values of V were mostly greater than 1 (open circles schematically in Fig. 7A), the pictures would be clearly distinguishable. If the values of V were generally less than 1 (squares in Fig. 7A), then the pictures should not be discriminable. Threshold might perhaps be achieved if the values of V fell on the dipper itself (filled circles). One very unlikely hypothesis (a "winner-take-all") is that discrimination would be possible provided that *just one* value $V_F(x,y)$ out of all the frequency bands and many locations exceeds the appropriate model dipper function.

A more realistic model would imagine that cues from different locations and frequency bands are pooled in some way, although the manner of pooling is debatable (Rohaly et al., 1997). We consider two possible pooling rules. First, we examine a rule in which the 68×68 V values in each frequency band are converted to logarithms (the y-axis of Fig. 7A is logarithmic) and averaged. The several frequency bands are kept separate, and discrimination will take place when this measure (*averaged-log*(V)) in any one of the frequency bands exceeds a certain "threshold" level. We will call this "*rule1*".

Figs. 7B and C show an implementation of this rule to the images constituting the *woman-to-man* sequence (b) and *car-to-bull* sequence (c) for observer CAP (foveal viewing of $\Delta\alpha = 0$ images). The graphs show, for five separate spatial frequency bands, how the "rule 1" *averaged-log*(V) increases as the test picture (abscissa) is made more different from the reference. The values for the *woman-to-man* sequence rise more slowly with per-

centage morph change than for the *car-to-bull* sequence, as might be expected given the degree of similarity between the faces and the dissimilarity between car and bull. An ordinate value of zero represents the point when the logarithms of the individual discriminability cues V are equally spaced above and below the dipper for that spatial frequency. Perhaps, the observer will be able to discriminate the pictures when the averaged cue just exceeds zero in at least one spatial frequency channel. This would be 12% morph in the *woman-to-man* sequence (frequency band 8 cycles/pic) and 4.5% for the bull-to-car sequence (frequency band 8 cycles/pic, although closely followed by 16 cycles/pic). These values are quite similar to the observer's actual thresholds for these stimuli (10.0% and 3.4%).

Fig. 8 shows a more detailed attempt to validate this pooling rule. In Fig. 8A, we plot the experimentally measured thresholds against those predicted by this simplistic model for 29 of the stimulus/observer combinations of the present experiments, for foveal viewing of the $\Delta\alpha = 0$ stimulus sets. The correlation between measured and predicted thresholds is convincing ($r = 0.57$, $n = 29$, $P = 0.001$). Fig. 8B shows additional foveal measurements with $\Delta\alpha = 0$ (see legend) mostly involving experiments on the good eyes of six amblyopic observers (Tolhurst & Párraga, 2003). Again there is a convincing correlation ($r = 0.64$, $n = 18$, $P < 0.01$). Figs. 8C and D show similar plots for all 28 threshold measurements (seven spectral slopes for each of four picture sets) for monocular foveal viewing by CAP and KB respectively. The correlation coefficients are 0.88 ($P < 0.001$) and 0.65 ($P < 0.001$).

The reasonable correlation coefficients suggest that a development of this simplistic model will be able to relate the thresholds for discriminating between complex natural scenes to the simple thresholds for detecting sinusoidal gratings. Fig. 8 also shows weighted least-squares regression lines, which allow the fit to be dominated by the data with the smallest standard errors. The slopes are close to unity and the intercepts close to zero (see figure legend). However, it is important to note that the regression lines are not the same as lines of equality, and there is much systematic scatter of the results about the regression lines. This shows that this model and/or pooling rule are too simplistic. Indeed, we tested the model with "pictures" of the sinusoidal gratings which were used to set up the model. We asked the model to compare gratings of a given contrast with ones whose higher contrast should have been just discriminable according to the theoretical dipper function. The *averaged-log*(V) value ranged from about -0.05 to -1.7 depending upon the spatial phase of the gratings, rather than being exactly zero. We will address this in Section 4.

A more likely hypothesis derives from the proposed probability summation in the *detection* of simple visual
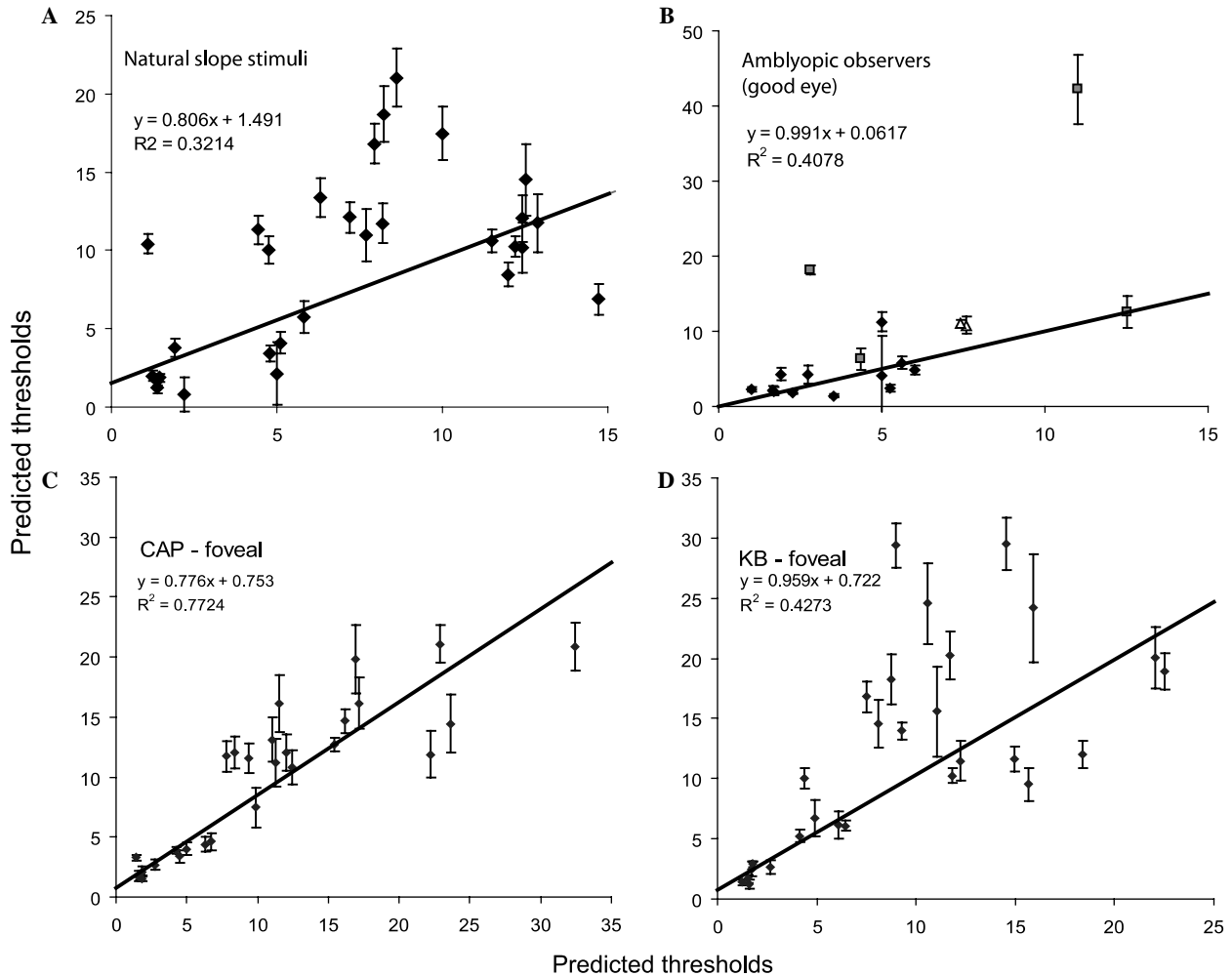
Fig. 8. (A) Discrimination thresholds for monocular foveal viewing of the $\Delta\alpha = 0$ stimuli in all four image series used in this paper (all observers). The experimentally measured threshold ($\pm$ one standard error) is plotted against the threshold predicted from the observers' CSFs using "rule 1a". 29 stimulus/observer combinations. The weighted least-squares regression line is shown which accounts for the different standard errors on the different measurements (slope = 0.81, intercept = 1.49%). (B) The circles show a similar analysis for six amblyopic observers using their good eye to view the car-to-bull and bull-to car series, while the triangles show predictions for two of the amblyopes viewing morph sequences of facial expressions (Tolhurst & Párraga, 2003); the squares show data for observers KB and CAP viewing a morphed sequence of a lemon turning into a pepper. Total data = 18. Weighted regression slope = 0.99, intercept = 0.062%. (C) For observer CAP, monocular foveal viewing, 28 measured and predicted thresholds are shown for each of 7 $\Delta\alpha$ amplitude spectrum slope increments and four different picture sets. Weighted regression slope = 0.78, intercept = 0.75%. (D) The same for observer KB. Weighted regression slope = 0.96, intercept = 0.72%.

stimuli (Graham & Robson, 1987; Graham, Robson, & Nachmias, 1978; King-Smith & Kulikowski, 1975; Robson & Graham, 1981; Tolhurst, 1975; Watson, 1979). For instance, the modelling in Fig. 7C suggests that *two* frequency bands might attain threshold together, and this would imply that the relevant images should be more discriminable than those in Fig. 7B where only one frequency band attains threshold. We use a weighted average of all the *V* cues, weighted across all locations and *all* frequency bands, so that there is a single metric for a given pair of pictures rather than one measure per frequency band. We use a Minkowski sum with power of 4 (Rohaly et al., 1997), and we call this "*rule 4*". The power of 4 derives from empirical

description of the amount of probability summation seen in grating *detection* experiments (see Section 4) and relates to the steepness of the psychometric function (Quick, 1974; Robson & Graham, 1981).

$$V_4 = \sqrt[4]{\sum_F \sum_x \sum_y (V_F(x, y))^4}. \tag{11}$$

We have no preconception of what value V_4 might take at threshold, but it was about 8.06 for pictures of gratings paired with ones of higher contrast that should have been just-discriminable according to the theoretical dipper function, *irrespective* of the spatial phase of the gratings.

3.6. Modelling the effects of spectral slope change and eccentricity

For each of observers KB and CAP, we measured their contrast sensitivity functions for gratings under the 4 foveal or eccentric viewing conditions, to produce 8 different versions of the discrimination model. We then examined whether the model was capable of explaining the forms of the experimental results: how threshold depends upon the amount of the increment or decrement in the slopes of the amplitude spectra of the images. Fig. 9 shows some examples of the procedures.

Figs. 9A and B shows how well the model fitted using "rule 1". The different lines show the model predictions presuming that discrimination will just be possible at different criterion values of *averaged-log(V)*. A single criterion value is used at a time to fit the seven experimental data points. Obviously, as the threshold criterion increases from −0.2 to 0 (Figs. 7B and C) to +0.3, so the predicted thresholds are higher; but the forms of the predicted curves also change. The model curves are "U"-shaped like the experimental values, and the "U" becomes sharper as the threshold criterion is raised. Although the models do not fit especially well, they *do* reflect that the observer's threshold curve *was* flatter fov-

eally than at 6°, that the observer's thresholds *were* lower foveally than peripherally, and that threshold *did* rise considerably at $\Delta\alpha$ of −1.2 peripherally. For each set of experimental results, we ran the model with both "rule 1" and "rule 4", and we adjusted the threshold criterion (*averaged-log(V)*, or V_4) to produce the best fit of the model, minimising the weighted residual sum of squares (SSE_R) between the model and the experimental results:

$$SSE_R = \sum_{i=1}^{n}[(Y_i - \hat{Y}_i)^2 \cdot W_i], \qquad (12)$$

where the weight W_i is the inverse of the squared standard error of each experimental measurement, and n is the number of data points fitted (7). This is essentially χ^2 (Press et al., 1986). We then defined an adjusted index of goodness-of fit, $\hat{\sigma}_A$, as

$$\hat{\sigma}_A^2 = \frac{SSE_R}{n-p} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \cdot W_i}{n-p}, \qquad (13)$$

where p is 1 when the model is allowed to search for the threshold criterion that minimises $\hat{\sigma}_A$, but is zero for "rule 1a" (see below) when the model is forced to use an *averaged-log(V)* threshold criterion of exactly zero. Seven points were considered in all cases ($n = 7$). When
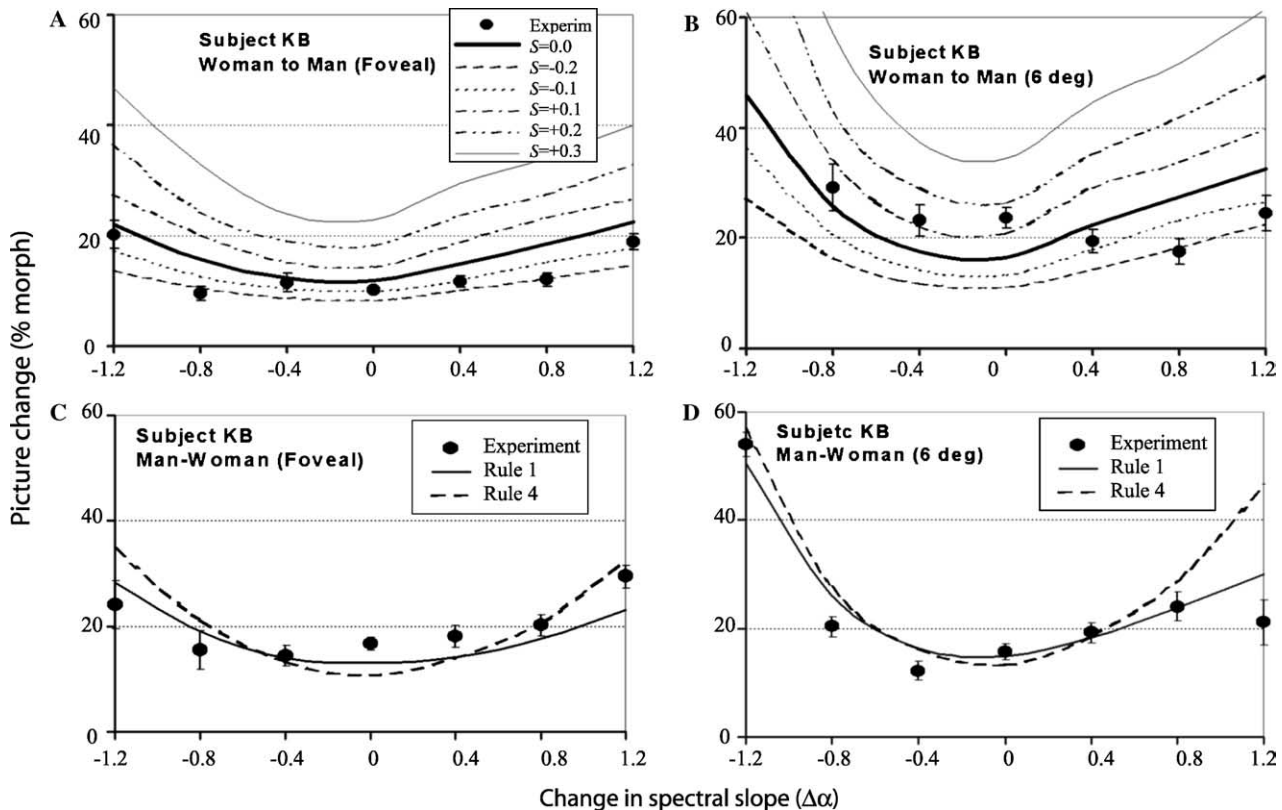

Fig. 9. (A) Fitting "rule 1" to the foveal results of KB *woman-to-man* images, for different values (−0.2 to +0.3) of the threshold criterion, *averaged-log(V)*. The circles show the experimental measurements with their standard errors. The thick line shows the model when the threshold criterion is zero ("rule 1a"). (B) The same, but for 6° peripheral viewing. (C) The best-fitting model curves for "rule 1b" (solid line) and "rule 4" (dashed line) for KB viewing the *man-to-woman* images foveally. (D) The same for 6° peripheral viewing.

the observer's measurements were missing (i.e., discrimination thresholds $\geqslant 100\%$ of the morph sequence), the thresholds were considered to be 100% and error bars of 50% were assumed. The same was applied to the model's results (e.g., when the model predicted thresholds $\geqslant 100\%$, they were considered to be equal to 100%). This allows us to take into account the cases when the model correctly predicted unmeasurably high discrimination thresholds. The smaller is $\hat{\sigma}_A$, the better the fit. Figs. 9C and D show further experimental results for observer KB, and the best fitting versions of the appropriate model under "rule 1" (solid lines) and "rule 4" (dashed lines). The two rules differ slightly in which aspects of the results they each best fit.

Ten more sets of experimental results are presented in Fig. 10, along with the corresponding model predictions for "rule 1" (solid lines) and "rule 4" (dashed lines). In fact, two versions of rule 1 are shown: the thick lines ("rule 1a") show how the model fared when it was constrained to use an *averaged-log*(V) criterion of zero (Fig. 8), while the thinner line allows the model to search for the *averaged-log*(V) criterion that minimises the weighted residual error ("rule 1b"). The selection of results in Fig. 10 is intended to show the range of behaviour of the results and the model, including two observers (CAP and KB), the bull/car sequences (the man/woman are illustrated in Fig. 9) and the four experimental viewing conditions (foveal, 3°, 6° and foveal-
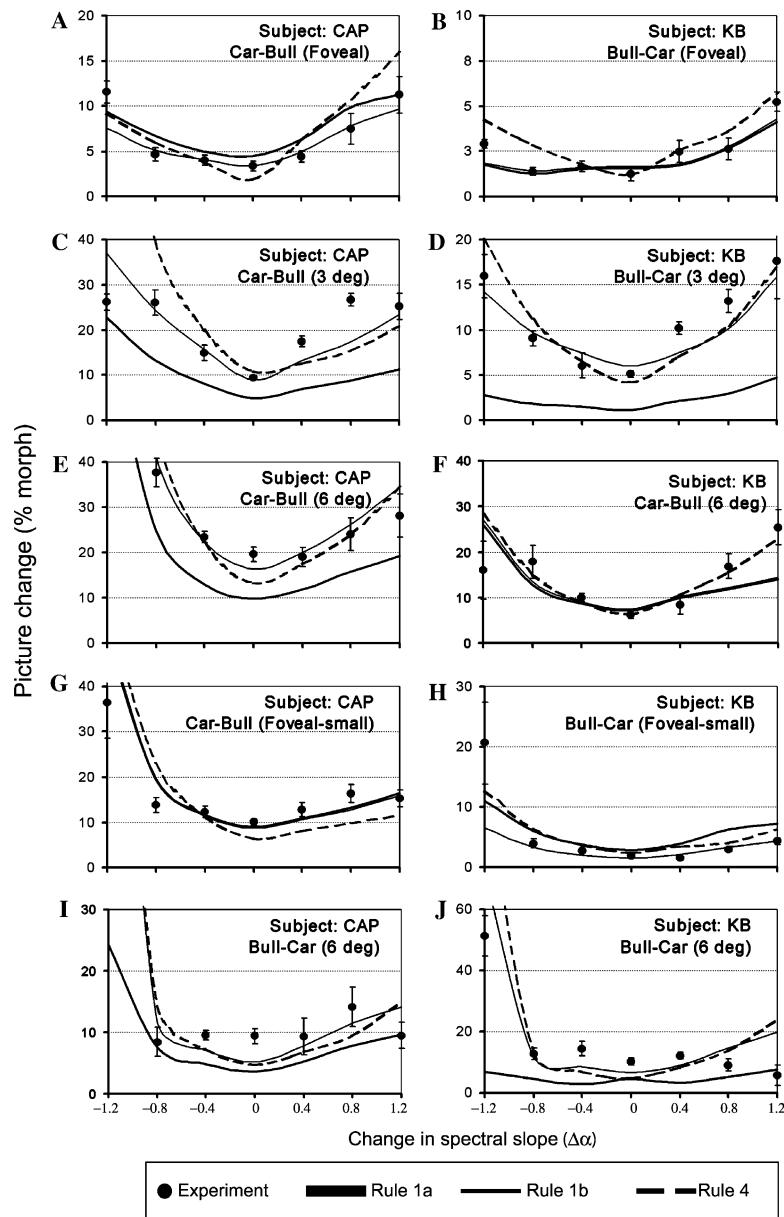

Fig. 10. Ten examples of model fits to the experimental results of two observers under four different viewing conditions. The circles show the experimental measurements with their standard errors. Three models are shown for each: "rule 1a" (thick solid lines), "rule 1b" (thin solid lines), and "rule 4" (dashed lines).

small pictures). In some cases, the models under "rule 1a" and "rule 1b" are very close (see panels A, B, G and H) but, in others, "rule 1a" overestimates the observer's performance and can produce a radically different form (e.g., panel J). Although individually some of the fits may not look very good, it is true that the model curves generally capture the different forms of the results: whether a particular graph is flattish, or is pronouncedly "U"-shaped, or whether it shows a sudden rise of threshold at $\Delta\alpha$ of -1.2 in the periphery. It is interesting that the models (Figs. 10E and F) are capable of explaining the differences in the results for the two observers at 6° eccentricity, which were shown in Figs. 4D and C, respectively. On the other hand, Figs. 10G and H show that the rise in threshold at $\Delta\alpha$ of -1.2 is captured by the model for only one of the observers.

Table 1 shows the values of the adjusted index of goodness-of-fit of the models to the monocular results for the 16 experimental conditions for observers KB and CAP. The Table shows that "rule 1b" provides the best fit (lowest values of $\hat{\sigma}_A$) in 43% of KB's experiments and 62% of CAP's, while "rule 4" comes second (with 31% of all experiments). The average values of the goodness-of-fit measure are 2.81 and 2.98 for "rule 1b" for KB and CAP respectively, and 3.22 and 4.67 for "rule 4". "Rule 1a" (with no degrees of freedom) comes surprisingly close (KB 5.22; CAP 4.22). This confirms the experience of Rohaly et al. (1997) that differences in the pooling rule in a multi-resolution model do not have a great effect. There seems no pattern as

to whether one rule might, perhaps, fit one eccentricity or one morphed image set better than another rule. The "rule 4" modelling of CAP's eccentric viewing results generally predicted that thresholds should be very high at $\Delta\alpha$ values of -1.2 and -0.8; thus, when CAP's thresholds were indeed high, the model gave a low σ_A but, when his thresholds were moderate, the σ_A values became unusually large.

Table 2 lists the values of the threshold criteria needed for best fit of the models for CAP and KB. The criterion values of *averaged-log*(V) needed for the best fit of "rule 1b" were 0.0249 for KB (average of 16 experiments, $SD = 0.213$) and 0.0670 for CAP (average of 16 experiments, $SD = 0.161$). The average values are close to our first naïve model which presumed that the threshold criterion might be zero, and might confirm the proposition that a complex image discrimination task *can* be interpreted in terms of simple grating thresholds; however, the standard deviations of these values are not small. The criterion values of V_4 needed for the best fit of "rule 4" were 40.99 for KB (average of 16, $SD = 11.96$) and 39.66 for CAP (average of 16, $SD = 11.65$). These values are substantially greater than that calculated for a grating at its contrast discrimination threshold (8.06).

Although the models explain the changes in *form* of the discrimination functions at different eccentricities, their success at explaining differences in threshold *magnitudes* is only partial. For a single picture set and a single observer, different values of V_4 were needed for the best fit of the model to the results for different viewing

Table 1
Adjusted index of goodness of fit ($\hat{\sigma}_A$)

	Rule 1a	Rule 1b	Rule 4	Rule 1a	Rule 1b	Rule 4
	KB—foveal			CAP—foveal		
Woman-to-man	3.25	1.52	2.66	3.60	2.08	3.13
Man-to-woman	4.69	1.41	2.51	2.32	2.44	3.87
Bull-to-car	2.11	2.14	3.78	2.86	2.96	6.16
Car-to-bull	3.99	4.08	4.78	2.09	1.45	2.34
	KB—3°			CAP—3°		
Woman-to-man	2.12	2.41	3.52	3.69	3.58	10.65
Man-to-woman	3.20	3.12	2.87	7.34	2.05	3.99
Bull-to-car	7.95	2.17	2.56	4.44	3.59	4.88
Car-to-bull	3.97	4.23	3.53	7.69	3.95	17.05
	KB—6°			CAP—6°		
Woman-to-man	2.56	2.84	3.96	4.27	3.73	5.66
Man-to-woman	3.27	1.97	3.30	3.28	2.14	1.55
Bull-to-car	5.06	3.15	4.79	3.11	2.21	2.54
Car-to-bull	1.78	1.81	1.10	4.46	1.13	2.04
	KB—foveal-small			CAP—foveal-small		
Woman-to-man	6.22	1.67	2.29	5.64	2.41	2.39
Man-to-woman	5.59	5.93	3.72	3.96	4.71	1.96
Bull-to-car	4.90	1.68	3.40	6.97	7.31	2.55
Car-to-bull	22.79	4.81	2.76	1.82	1.99	3.89

A summary of the modelling of the experimental results for observers KB and CAP, who each performed experiments under four foveal or eccentric conditions and who performed experiments with four different stimulus sets under each condition. The adjusted index of goodness of fit ($\hat{\sigma}_A$) is defined in Eqs. (10) and (11). A small value means a good fit. Some of the "rule 1b" values are *higher* than the corresponding "rule 1a" values, even though "rule 1b" has an extra fitting parameter that should lead to a better fit—note that σ_A is divided by the *remaining* degrees of freedom. The three pooling rules are described in the text.

Table 2
Threshold criterion for best fit of model

	Rule 1b *averaged-log(V)*	Rule 4 V_4	Rule 1b *averaged-log(V)*	Rule 4 V_4
	KB-foveal		CAP-foveal	
Man-to-woman	0.255	53.69	0.0245	39.29
Woman-to-man	−0.126	29.72	−0.113	31.47
Bull-to-car	0.0133	48.30	0.0099	46.62
Car-to-bull	0.133	46.91	−0.0976	38.50
	KB—3°		CAP—3°	
Man-to-woman	0.0449	42.94	0.4227	57.66
Woman-to-man	−0.039	33.58	0.1288	37.73
Bull-to-car	0.436	68.66	0.1346	43.60
Car-to-bull	0.0299	40.58	0.2369	62.06
	KB—6°		CAP—6°	
Man-to-woman	0.0894	41.61	0.117	37.85
Woman-to-man	−0.084	28.76	−0.0884	22.83
Bull-to-car	0.3799	57.78	0.152	40.74
Car-to-bull	0.0149	40.10	0.242	43.36
	KB—small-foveal		CAP—small-foveal	
Man-to-woman	0.0491	35.03	−0.125	33.32
Woman-to-man	−0.187	22.49	−0.134	23.12
Bull-to-car	−0.227	33.46	0.173	34.07
Car-to-bull	−0.384	32.16	−0.0123	29.94

A summary of the modelling of the experimental results for observers KB and CAP, who each performed experiments under four foveal or eccentric conditions and who performed experiments with four different stimulus sets under each condition. The values of the threshold-criterion needed to best the fit the model to the experimental results under rules 1b and 4 are shown.

eccentricities. The different magnitudes of the thresholds, as predicted by the model, depend partly on differences in the observer's CSF at different eccentricities, but also on unexplained differences in the criterion V_4 needed for the best fit. For instance, for observer CAP viewing the *car-to-bull* images, the criterion V_4 needed for best fit was 62.06 for 3° viewing (Fig. 10C, dashed line), but only 29.94 for "small-foveal" viewing (Fig. 10G); the criterion values are different, even though the observer's thresholds at Δα of zero are similar (9.47 and 10.05%). Here, the success of the model's fit seems to rely on a large difference in a parameter (V_4) whose values we are still unable to explain. On the other hand, CAP's results for foveal viewing (Fig. 10A, dashed line) and 6° viewing (Fig. 10E) *were* fit with very similar values of V_4 (38.50 and 43.36, respectively), even though the observer's thresholds at Δα of zero differed markedly: 3.40% foveally and 19.64% peripherally. In this case, at least, the very different forms and magnitudes of the thresholds at the two viewing conditions *are* explained by differences in the observer's grating CSF for the two conditions.

The criterion values of V_4 for the two observers are very similar on average and are correlated ($r = 0.54$; $n = 16$; $P = 0.032$). This is most interesting for 6° eccentric viewing of the *car-to-bull* images (Figs. 4C and D—filled circles, and Figs. 10E and F—dashed lines) where the observers' thresholds were very different but the V_4 values were almost the same (KB 40.10; CAP 43.36). The differences in the observers' thresholds in this experiment must have been due to the differences in their 6°

CSFs (CAP's sensitivity was 5–7 times lower than KB's at frequencies above about 10 cycles/pic). The "rule 4" V_4 values were highly correlated with the "rule 1b" *averaged-log(V)* values (KB $r = 0.88$; CAP $r = 0.68$) although the *averaged-log(V)* values for the two observers were poorly correlated ($r = 0.24$). The significant correlations imply that the model behaves in consistently different ways for some image/eccentricity combinations. Indeed, the *woman-to-man* images required substantially lower values of V_4 (KB 28.64; CAP 28.11) than the other three image sets, and the small-foveal viewing condition needed smaller V_4 values (KB 30.79; CAP 30.11) than the other three viewing conditions (compare Ripamonti, Tolhurst, Lovell, & Troscianko, 2005).

3.6.1. A single-channel model

We also modelled CAP's foveal data with a single-channel model; the single circularly-symmetric channel had a CSF identical to that of the observer. We used "rule 1b" and "rule 4" and all four morph sequences. For three of the image sequences, the single channel model produced a definitely worse fit of the data, with GoF value between 13% and 200% higher. Only in the bull-to-car sequences, did the single channel version of the model seem to fit the data better than the multiple-channel model, giving some 50% better GoF values on average. In general, the single channel model failed to explain the large rise in threshold found for negative values of Δα ("whitened" images).

4. Discussion

We have measured thresholds for discriminating small morphed spatial changes in naturalistic stimuli, and we have examined how those thresholds are affected when we distort the images' amplitude spectra from their natural state. In almost all individual experiments, the observers were best able to discriminate the morphed changes when the amplitude spectra were close to having "natural statistics"; in fact, there was a broad threshold minimum consistent, perhaps, with the wide range of naturally occurring spectral slopes (Tolhurst et al., 1992). This is similar to the results of (Tolhurst & Tadmor, 2000) who used spectrally blended natural images rather than morphed ones.

To make morphed sequences, it is necessary to begin with parent images of well defined objects (preferably against a blank background). Such images can be natural (view an object on a hill against a blank sky) but their amplitude spectra fall at one extreme of the natural range: the slopes of the spectra of our parent images were steep. The blending technique of (Tolhurst & Tadmor, 2000) is not so constrained, and their similar results were obtained with parent images much more similar to the average values for natural images. On the other hand, the blending technique makes intermediate images which may have naturalistic statistics but which are not representations of everyday natural objects or scenes—the blended images have "ghosts" of the parent images. Every image in the morphed sequence is potentially realisable as an object, though we admit that a hybrid car/bull is not likely; in practice, the observers' threshold for the car/bull sequences were very small, so that the observers *were* looking at subtle differences in bulls or subtle differences in cars.

That the observers were able to perform the discrimination tasks best when the images had near-natural amplitude spectra might be taken as experimental evidence for the popular contention (Barlow, 1961; Laughlin, 1983; Marr, 1982) that the visual system is optimised for processing stimuli with natural statistics. Although very influential, this hypothesis is supported by only a little psychophysical evidence that vision is actually "best" in any sense with natural stimuli (Geisler et al., 2001; Knill et al., 1990; Tadmor & Tolhurst, 1994; Tolhurst & Tadmor, 2000) . Although there is much theoretical evidence that information encoding might be most efficient when images have natural statistics, it is not necessarily the case that processing of non-natural statistics will therefore be inefficient. Furthermore, any relative inefficiency in encoding of non-natural scenes might not be reflected as an elevation of a simple discrimination threshold. "Inefficiency" might be exhibited in something very difficult to measure, such as removal of neural resources from some other (perhaps non-visual) function or an increased amount of metabolic energy

needed for a task! However, we would be surprised if efficient encoding at low-levels of the visual system did not confer some advantage in visual performance.

Unfortunately, the changes that we have imposed on the amplitude spectra to confer different degrees of unnaturalness do change the overall visibility and contrast of the images. Changes in image contrast do not grossly change the form of the results (Párraga et al., 2000), and our analysis of the RMS pixel differences between images at threshold (Fig. 3) suggests that the observers are not using some simple image metric for detection. However, it may be that the characteristic "U"-shaped graphs we present are influenced by relatively low-level changes in visibility of some spatial-frequency bands. An experimental "proof" of the hypothesis that vision is optimised for natural statistics might try to distort the image statistics in a way that does not immediately compromise visibility. For instance, one might try to distort the "higher-order" statistics characteristic of natural images (e.g., Geisler et al., 2001; Thomson & Foster, 1997) rather than the lower, second-order (amplitude-spectra) statistics (Knill et al., 1990).

We might be able to argue that our results imply that foveal vision (see Fig. 5, open symbols) is optimised for natural image statistics, since thresholds are lowest when the amplitude spectra are undistorted (Tolhurst & Tadmor, 2000). However, this same argument then raises the question why peripheral vision should *not* be similarly optimised: at 6° eccentricity, thresholds are minimal for images that are blurred compared to natural ones (Fig. 5 grey symbols). This shift in the minimum was matched by M-scaling the foveal stimuli to match the supposed cortical magnification at 6° eccentricity. The thresholds for making the morph discrimination were generally higher in the periphery than in the fovea, but this was a result that could *not* be replicated by M-scaling of the foveal stimuli; M-scaling did raise foveal thresholds, but not nearly enough to match the high peripheral ones. Perhaps, our choice of M-scaling factor was too conservative (see Tolhurst & Ling, 1988), reflecting simple acuity tasks rather than tasks requiring more neural processing. We estimated the values of S (inverse of the eccentricity at which the task becomes twice as difficult as in the fovea) for our discrimination tasks. Our results show that $S = 0.12 \deg^{-1}$ for the man/woman sequences and $S = 0.43 \deg^{-1}$ for the car/bull sequences. These values of S are of the same order of magnitude as those measured for grating acuity tasks (and 0.38 in Klein & Levi, 1987; 0.41 in Virsu, Näsänen, & Osmoviita, 1987).

4.1. A model of visual discrimination

We have been developing a relatively simple multiple spatial-frequency channel model of the low-level

discrimination process, similar to other algorithms such as *DCTune* (Watson, 1983; Watson, 1987; Watson, 1993), *VDP* (Daly, 1992; Lubin, 1993) and the model of Rohaly et al. (1997). Our implementation has previously had some success in modelling the detectability of changes in the amplitude spectra of natural scenes (Tolhurst & Tadmor, 1997). Discrimination models are needed for several reasons (e.g., Ferwerda & Pellacini, 2003; Mitchell et al., 2000). First, they can help assess the quality of image displays, in which veridical representation of scenes may be necessary (e.g., in aeroplane pilots' training, quality control, surveillance, etc.). Second, and conversely, they can help assess the quality of simulated natural images, to avoid excessive rendering which would be unlikely to make any real extra impact on image quality, thus saving time, bandwidth and processing power. Third, they provide important clues about the function of visual mechanisms; a model based closely on, say, visual cortex neuronal properties would allow us to evaluate whether our immense knowledge of such neurons is actually adequate to explain vision in the real world.

Our model is based on evidence, both physiological (De Valois et al., 1982; Movshon et al., 1978) and psychophysical (Blakemore & Campbell, 1969; Campbell & Robson, 1968; Legge & Foley, 1980), for the existence of multiple channels tuned to spatial frequency. The spatial contrast sensitivity function (the overall CSF) is presumed to be the envelope of many narrowly tuned frequency selective channels. In the model, the differences in contrast (Peli, 1990; Tadmor & Tolhurst, 1994) between two images are calculated within a number of spatial-frequency channels designed to have the same spatial-frequency bandwidth as simple cells in the visual cortex (about 1–1.5 octaves). We presume that simple cells in several independent spatial-frequency bands sample the reference and test stimuli point-by-point, and that each cell then signals any local differences in the spatial structure of the two stimuli.

Each cell contributes some cue to the overall discrimination, and the size of a difference cue is determined from the "dipper" function for discriminating between contrasts of sinusoidal grating. The cues from the many cells must be combined to give an overall indication of how discriminable two images are. We have considered two ways in which these many cues might be pooled. In "rule 1", we presumed that discrimination would just be possible when the average of the individual cues (expressed as logarithms above or below the dipper) reached some criterion value. This is rather simplistic, but it does lead to a straightforward prediction about the magnitudes of the discrimination thresholds in the complex scenes (see below). "Rule 1b" is a variant which allows the model to search for the threshold criterion that minimises the goodness of fit. We also used Minkowski summation with an exponent of 4 (Quick,

1974; Watson, 1987) by analogy with models of probability summation in the *detection* thresholds for sinusoidal gratings. In fact, psychometric functions may be shallower for discrimination than for detection (Bird, Henning, & Wichmann, 2002; Chirimuuta & Tolhurst, in press) but the choice of a pooling rule appeared to have surprisingly little effect on the goodness of the model fits to the experimental results (Table 1), as also found by Rohaly et al. when they changed the magnitude of the Minkowski exponent (Rohaly et al., 1997). In "rule 1", all cues count towards the average, even those that are miniscule; in "rule 4", discrimination is determined by a subset of cells, those giving the largest cues.

We "customise" the model to include each observer's CSF for sinusoidal gratings in the different foveal and peripheral viewing conditions. In most cases, the customised model is able to explain the overall "U"-shaped form of the results, including the finding that the thresholds for the car/bull sequences are lower than for the man/woman sequences. When we customise the model to use the observers' peripheral CSFs for sinusoidal gratings rather than the foveal CSF, the model generally explains that the thresholds for making morphed discriminations are higher in periphery than in the fovea, but *not* that the minimum of the "U" is shifted for 6° peripheral viewing. When we customise by using the CSFs of different observers, the model can explain some of the differences in the form of the results between different observers.

Thus, it is fundamental to the model that the magnitudes of the thresholds for complex, natural-image discriminations will depend upon the magnitudes of the contrast-detection thresholds for sinusoidal gratings. In our experience, such modelling is capable of predicting thresholds for one stimulus or one observer *relative* to another; it is harder to make an *absolute* prediction. To move the relative thresholds into the correct absolute range, we lose a degree of freedom to a free parameter—the "threshold criterion" (Table 2). However, for the pooling "rule 1a", we *were* able to make a simplistic assumption without losing this degree of freedom and, thence, an absolute prediction of the thresholds for morph discriminations. Fig. 8 shows that this naive model of pooling is remarkably good at predicting the absolute magnitudes of morph discrimination thresholds from grating detection thresholds. In general, "rule 1a" performed creditably (Table 1) compared to "rule 1b" and "rule 4", despite its having one fewer degrees of freedom.

4.2. The threshold criterion values given by the model

Consider that we have two images which are just different enough to be at psychophysical discrimination threshold. These might be two images from our

morphed sequences, or two gratings of slightly different contrast. The model compares the two paired images and returns with one or more "numbers" that ideally represent the perceptual difference between the images. We would presume that any pair of images, if they were just at threshold, would return the same values—the threshold criterion. However, this was not the case.

First, the threshold criterion values calculated for sinusoidal gratings of different contrast were very different from the values calculated for pairs of naturalistic images. In "rule 1", where the visibility cues are averaged across the whole image, the threshold criterion changed dramatically with the spatial phase of the grating. This was an arithmetic artefact; as spatial phase changed, so the miniscule visibility cues near the zero-crossings of the grating changed. They may have changed several orders of magnitude, but were always tiny; their role in the arithmetic was significant because they were expressed as logarithms. "Rule 4" relies only on cells giving the largest cues, and so these artefactual differences in tiny cues were not a problem. However, the stable "rule 4" criterion values for grating discrimination were much less than those calculated for the morphed images. The model suggested that, compared to grating discrimination, our ability to discriminate between morphs was poor.

Second, different criterion values were needed to explain the discrimination thresholds for the different sequences of morphed images and the different eccentricity viewing conditions. The values needed for the two modelled observers were similar and correlated (see Table 2). The "rule 4" criterion value for the *man-to-woman* sequence was 42.67 (averaged across four viewing conditions and the two observers). This was very similar to the averaged criterion value for the *bull-to-car* (46.65) and for the *car-to-bull* (41.70) sequences, and this similarity represents a success of the model since the discrimination thresholds for the *man-to-woman* morphs were very different from those in the two car/bull sequences. However, the failure of the modelling is also illustrated: the averaged criterion value for the *woman-to-man* sequences was substantially lower (28.71), even though the *woman-to-man* thresholds were almost the same as those for the *man-to-woman* sequence.

There are, thus, consistent failures in the detailed implementation of the discrimination model both for naturalistic images and for gratings. The big inconsistency between the modelling of grating discrimination and morph discrimination implies that we have not correctly modelled the differences between narrow-band and broad-band stimuli. Perhaps, probability summation does not work as uniformly across spatial location and spatial-frequency scale as we have modelled. For instance, our natural scene stimuli are likely to have multiple, spatially separated cues to discrimination, and an observer may not be able to locate or attend to all of them. Or, there may be different kinds of cue such as changes in object shape, contrast and texture; the model will detect all of these cues but, for some reason, the observer may fail to perceive some of them. We also ignore any effects of eye movements in our relatively small, briefly presented, pictorial stimuli.

4.3. Future development of the discrimination model

We have experimented with, e.g., changing the bandwidth of the contrast operators, the position and form of the CSF, the value of the Minkowski exponent and the form of the dipper function, but have had no systematic improvement in the way that the model fits experimental data or any resolution of the inconsistencies in threshold criterion values. Indeed, the model is rather tolerant of detailed changes, as is shown by the fact that a single-channel model is only slightly less effective than the multiple-channel model. To obtain an insight into the workings of the model and the effects of its various components (bandwidth of the channels, shape and position of the "dipper" function and the CSF), we explored how the model's predictions changed as the model parameters were altered. For simplicity, we chose to test only the least complex version of the model (rule 1a) applied only to the woman-to-man sequence using observer CAP's foveal CSF (Párraga et al., 1999). Changes in the spatial frequency bandwidth of the channels from 0.5 to 1.5 octaves produced relatively little change in the predicted thresholds, while bandwidths of 1.9 octaves and more produced higher thresholds especially for "whitened" ($\Delta\alpha < 0$) sequences, with adjusted goodness of fit (GoF) indices some 200% higher than in the 1 octave case. A "dipper" function with unity slope (consistent with Weber's Law) produced higher predicted thresholds on the "blurred" side ($\Delta\alpha > 0$) leading to a GoF coefficient some 300% higher than those obtained with the "biological" dipper shown in Fig. 7A. Altering the shape of the observer's CSF (by increasing the model's sensitivity to higher spatial frequencies) did have the effect of over-predicting thresholds in the "whitened" side of the plot, presumably because high spatial frequencies do play a dominant role there. However, using a "flat" CSF with the same sensitivity at all frequencies made the GoF coefficients only 30% higher. Shifting the observer's CSF down three times (lower sensitivity) produced a GoF coefficient about 300% higher. Making the opposite change only produced a marginal improvement in GoF. In summary, no single parameter seems to be responsible for the GoF.

Presumably, more features of the experimental results could be explained more reliably if the model were better matched to the known behaviour of real V1 neurons. Indeed, we have ignored orientation tuning (Campbell & Kulikowski, 1966; Hubel & Wiesel, 1959) which is

one of the most obvious features of channels and V1 neurons; however, we have implemented orientation tuning (unpublished observations) to model other experiments (e.g., Ripamonti et al., 2005) and we have yet to find a situation where its inclusion leads to different conclusions. Nor have we yet modelled contrast normalisation or non-specific suppression (Foley, 1994; Heeger, 1992; Watson & Solomon, 1997) which Rohaly et al. (1997) suggest is crucial to successful modelling of broad-band naturalistic stimuli. Such normalisation would reduce both the magnitude of the "pedestal" and of the "increment" by the same factors in the putative contrast discriminations which we model as underlying the discrimination between morphed images. However, since the "Weber" part of the dipper function does not have unity slope on log–log co-ordinates, even proportionate changes in effective contrast would affect discrimination. Furthermore, we would expect the form of the dipper to be changed by normalisation (Chirimuuta & Tolhurst, in press; Foley, 1994).

The present model clearly needs refinement but, even in its present simplistic form, it is capable of explaining many of the differences in thresholds between observers, between eccentricities and between picture sets. The latter is important, since it gives us confidence that experiments on a few examples of natural images may be representative of a much wider array. Morphing allows us to generate images that could represent real faces or real objects, but only certain sorts of image can be morphed; all of our present images are based around single portraits/objects filling the bulk of the stimulus area, seen against a uniform background. Perhaps, with experiments and modelling on a wider variety of stimulus images, we will be able to trace and correct the inconsistencies in the present model.

Acknowledgments

References

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communications* (pp. 217–234). Cambridge, MA: MIT Press.

Benson, P. J. (1994). Morph transformations of the facial image. *Image and Vision Computing, 12*, 691–696.

Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *Journal of the Optical Society of America A, 19*, 1267–1273.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237–260.

Brady, N., & Field, D. J. (2000). Local contrast in natural images: normalisation and coding efficiency. *Perception, 29*, 1041–1055.

Burton, G. J., & Moorhead, I. R. (1987). Color and Spatial Structure in Natural Scenes. *Applied Optics, 26*, 157–170.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197*, 551–566.

Campbell, N. W., & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *Journal of Physiology, 187*, 437–445.

Carlson, C. R. (1978). Thresholds for perceived image sharpness. *Photographic Science and Engineering, 22*, 69–71.

Chirimuuta, M. & Tolhurst, D. J. (in press). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, doi:10.1016/j.visres.2005.06.022.

Clatworthy, P. L., Chirimuuta, M., Lauritzen, J. S., & Tolhurst, D. J. (2003). Coding of the contrasts in natural images by populations of neurons in primary visual cortex (V1). *Vision Research, 43*, 1983–2001.

Daly, S. (1992). The visible differences predictor: An algorithm fo the assessment of image fidelity. In *Proceedings SPIE the International Society For Optical Engineering, SPIE—Society of Photo Optical Instrumentation Engineers* (Vol. 1666, pp. 2–15).

Daly, S. (1993). The visible differences predictor: an algorithm for the assesment of image fidelity. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 179–206). Cambridge, MA: MIT Press.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial-frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Doll, T. J., McWorter, S. W., Wasilewski, A. A., & Schmieder, D. E. (1998). Robust, sensor-independent target detection and recognition based on computational models of human vision. *Optical Engineering, 37*, 2006–2021.

Drasdo, N. (1991). Neural substrates and threshold gradients of peripheral vision. In I. J. Murray (Ed.). *Limits of vision* (5, pp. 251–276). London: Macmillan Press Ltd.

Edwards, A. W. F. (1972). *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*. London: Cambridge University Press.

Ferwerda, J. A., Pellacini, F. (2003). Functional difference predictors (FDPs): Measuring meaningful image differences. In *37th asilomar conference on signals, systems and computers, Asilomar Hotel and Conference Grounds Pacific Grove, CA* (Vol. 9137, pp. 1388–1392), IEEE.

Field, D. J. (1987). Relations between the statistics of natural scenes and the response properties of cortical-cells. *Journal of the Optical Society of America A, 4*, 2379–2394.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation, 6*, 559–601.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A, 11*, 1710.

Garcia-Perez, M. A., & Peli, E. (2001). Luminance artifacts of cathode-ray tube displays for vision research. *Spatial Vision, 14*(2), 201–215.

Geisler, W. S. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 825–838). Cambridge, MA: MIT Press.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711–724.

Graham, N., & Robson, J. G. (1987). Summation of Very Close Spatial-Frequencies - the Importance of Spatial Probability Summation. *Vision Research, 27*(11), 1997–2007.

Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research, 18*, 815–825.

Hancock, P. J. B., Baddeley, R. J., & Smith, L. S. (1992). The principal components of natural images. *Network-Computation in Neural Systems, 3*, 61–70.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Kersten, D. (1987). Predictability and redundancy of natural images. *Journal of the Optical Society of America A, 4*, 2395–2400.

King-Smith, P. E., & Kulikowski, J. J. (1975). Pattern and flicker detection analyzed by subthreshold summation. *Journal of Physiology, 249*, 519–548.

Klein, S. A., Hu, Q. J., & Carney, T. (1996). The adjacent pixel nonlinearity: Problems and solutions. *Vision Research, 36*, 3167–3181.

Klein, S. A., & Levi, D. M. (1987). Position sense of the peripheral retina. *Journal of the Optical Society of America A, 4*, 1543–1553.

Knill, D. C., Field, D., & Kersten, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America A, 7*, 1113–1123.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung C, 36*, 910–912.

Laughlin, S. B. (1983). Matching coding to scenes to enhance efficiency. In Rank Prize Funds (Ed.), *Physical and biological processing of images: proceedings of an international symposium organized by the Rank Prize Funds, London, England, 27–29 September, 1982* (pp. 42–52). Berlin; New York: Springer-Verlag.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research, 21*, 457–467.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*, 1456–1471.

Levi, D. M., Klein, S. A., & Aistebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*, 963–977.

Lubin, J. (1993). The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 163–178). Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Menendez, A. R., & Peli, E. (1995). Vision models for target detection and recognition: in memory of Arthur Menendez. *Series on information display* (Vol. 2). Singapore, River Edge, NJ: World Scientific.

Mitchell, K. D., Moorhead, I. R., Gilmore, M. A., Watson, G. H., Thomson, M., Yates, T., Troscianko, T., & Tolhurst, D. J. (2000). Assessment of synthetic image fidelity (4029-30). In *Proceedings SPIE the International Society For Optical Engineering, SPIE—Society of Photo Optical Instrumentation Engineers* (pp. 256–266).

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology, 283*, 101–120.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast discrimination may be better than detection. *Vision Research, 14*, 1039–1042.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 3311–3325.

Osorio, D., & Vorobyev, M. (1996). Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London Series B, 263*, 593–599.

Párraga, C. A., Brelstaff, G., Troscianko, T., & Moorhead, I. R. (1998). Color and luminance information in natural scenes. *Journal of the Optical Society of America A, 15*, 563–569.

Párraga, C. A., & Tolhurst, D. J. (2000). The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes. *Perception, 29*, 1101–1116.

Párraga, C. A., Tolhurst, D. J., & Troscianko, T. (1999). A computational model predicts discrimination thresholds for morphed objects in natural scenes. *Perception, 28*(Suppl.), 127b.

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

Párraga, C. A., Troscianko, T., & Tolhurst, D. J. (2002). Spatiochromatic properties of natural images and human vision. *Current Biology, 12*, 483–487.

Párraga, C. A., Troscianko, T., Tolhurst, D. J., & Gilchrist, I. D. (2000). Discrimination thresholds for morphed objects in peripheral vision. *Perception, 29*(Suppl.), 58.

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A, 7*, 2032–2040.

Pelli, D. G. (1997). Pixel independence: Measuring spatial interactions on a CRT display. *Spatial Vision, 10*, 443–446.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contast on microcomputer displays. *Vision Research, 31*, 1337–1350.

Press, W. H., Flannery, B. P., Tekulosky, S. A., & Vetterling, W. T. (1986). *Numerical recipes*. New York: Cambridge University Press.

Quick, R. F. (1974). A vector magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Regan, B. C., Julliot, C., Simmen, B., Vienot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions Royal Society of London Series B Biological Sciences, 1407*, 229–284.

Ripamonti, C., Tolhurst, D. J., Lovell, G., Troscianko, T. (2005). Magnification factors in a V1 model of natural-image discrimination. In *Vision Sciences Society fifth annual meeting, E94 597, Sarasota, FL, USA*.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21*, 409–418.

Rohaly, A. M., Ahumada, A. J., & Watson, A. B. (1997). Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research, 37*, 3225–3235.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research, 37*, 495–510.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Research, 39*, 2697–2716.

Schofield, A. J., & Georgeson, M. A. (2000). The temporal properties of first- and second-order vision. *Vision Research, 40*, 2475–2487.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In N. N. Osborne & G. J. Chader (Eds.). *Progress in Retinal Research* (3, pp. 263–346). Oxford: Pergamon Press.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding—a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London Series B, 216*, 427–459.

Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second-order statistics of natural and synthetic-images. *Vision Research, 34*, 541–554.

Tadmor, Y., & Tolhurst, D. J. (2000). Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Research, 40*, 3145–3157.

Thomson, M. G. A., & Foster, D. H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A, 14*, 2081–2090.

Tolhurst, D. J. (1975). Sustained and transient channels in human vision. *Vision Research, 15*, 1151–1155.

Tolhurst, D. J., & Ling, L. (1988). Magnification factors and the organisation of the human striate cortex. *Human Neurobiology, 6*, 247–254.

Tolhurst, D. J., & Párraga, C. A. (2003). Visual discrimination of natural-scene stimuli by amblyopic people. *Journal of Physiology, 548P*, O75.

Tolhurst, D. J., & Tadmor, Y. (1997). Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra. *Vision Research, 37*, 3203–3215.

Tolhurst, D. J., & Tadmor, Y. (2000). Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images. *Perception, 29*, 1087–1100.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12*, 229–232.

Tolhurst, D. J., & Thompson, I. D. (1981). On the variety of spatial-frequency selectivities shown by neurons in area-17 of the cat. *Proceedings of the Royal Society of London Series B, 213*, 183–199.

Van Hateren, J. H., & Van Der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London Series B, 265*, 359–366.

Virsu, V., Näsänen, R., & Osmoviita, K. (1987). Cortical magnification and peripheral vision. *Journal of the Optical Society of America, 4*, 1568–1578.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*, 515–522.

Watson, A. B. (1983). Detection and recognition of simple spatial forms. In O. J. Braddick (Ed.). *Physical and biological processing of images: proceedings of an international symposium organized by the Rank Prize Funds, London, England, 27–29 September, 1982* (Vol. 11, pp. 100–114). Berlin, New York: Springer-Verlag.

Watson, A. B. (1987). Efficiency of a model human image code. *Journal of the Optical Society of America A, 4*, 2401–2417.

Watson, A. B. (1993). DCTune: A technique for visual optimization of DCT quantization matrices for individual images. *Sid International Symposium Digest of Technical Papers, 24*, 946.

Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optical Express, 6*, 12–33.

Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research, 21*, 1115–1122.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A, 14*, 2379–2391.