Brief Communication:

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

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

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

Running head: Vision is optimised for natural images.

Corresponding author: C.A. Párraga (Alej.Parraga@bris.ac.uk)

Keywords: Vision; Visual cortex; Natural images; Spatial vision; Power spectra;

Abstract

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

Results and Discussion

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

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

Where f is spatial frequency and α (the slope constant) ranges from about 0.7 to about 1.6[18-20] (Fig.2). Thus, we can make a set of pictures more or less unnatural in a *systematic* way by decreasing (Fig.1b) or increasing (Fig.1c) the slopes of their amplitude spectra. If there *is* an optimum in performance when the second-order statistics correspond to those of natural scenes, we will be able to conclude that those statistics have been important in shaping the function of our visual system.

Figures 1 and 2 near here

Figure 3 shows the experimental results for two observers, CAP and TT, who are authors of this paper. The main effects were also replicated on a psychophysically experienced but naïve observer. Each observer was presented with four types of reference stimulus: pictures of a bull, a car, a man's face and a woman's face. The observer had to identify test pictures that were different from the reference, in the direction of the other end of the morph series (see Fig.1). Thus, a car would become slightly more like the bull, the bull more like the car, the man's face more like the woman's, and the woman's face more like the man's.

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

Our prediction is that, if "natural" scenes are optimally encoded, the thresholds at zero amplitude slope offset will be lower than the other values. And, indeed, the 8 experimental sets do generally show that threshold is lowest between amplitude slope offsets ($\Delta\alpha$) of –0.4 and +0.4, and that threshold rises more or less symmetrically from this lowest point. We can show this more formally by fitting two linear equations to each dataset and looking at where they intersect, or by fitting second-order polynomials. The mean intersection point for pairs of lines was at a slope offset $\Delta\alpha$ = -0.071 (standard deviation 0.425). The mean slope offset of the minima of the fitted polynomials was $\Delta\alpha$ = -0.016 (standard deviation 0.175). χ^2 of these fits were compared to the χ^2 resulting from fitting a single line; they were 13 and 6 times smaller on average, respectively.

The solid and dashed curves are predictions of the thresholds for each observer and for each morphed sequence made by a model of local contrast discrimination between the reference picture and test pictures progressively differing from it in the morph series. This model has as its input each observer's contrast sensitivity function (measured with sinusoidal gratings for this purpose). It locally decomposes each scene into its constituent spatial frequency components, and measures the energy in spatial-frequency bands of amplitude 1.5 octaves[15-17]. It then compares the energy of each filter in the two images, and computes whether this lies above discrimination threshold as measured by the standard contrast discrimination function of, e.g., Legge and Foley[21]. The predicted threshold is obtained by computing that morphed image which has just supra-threshold discriminability in at least one spatial-frequency band. The fits of these predicted curves to the data are reasonably good, particularly considering that the inputs to the model are only the detectability and discriminability of sinusoidal gratings while, in this task, observers were discriminating complex images. There are *no* free parameters to the model, and the curves have not been shifted up or down the y-axes to enhance the fit.

Figure 3 near here

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

Figure 4 near here

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

Conclusions

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

Methods

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

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

$$Weight(f) \propto f^{-\Delta\alpha} \qquad \text{Eqn.2}$$

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

Experiments: Thresholds were measured using a modified two-alternative forced choice paradigm. The observer was presented with three pictures sequentially; each was presented for 500 ms with intervals of 200 ms between. The second presenetation always contained a copy of the reference picture; one of the other two presentations (chosen randomly by computer) contained an identical copy of the reference, while the remaining presentation contained a morphed picture. The observer had to press mouse buttons to tell the controlling computer whether the morphed picture was the first or the last of the sequence of three. Auditory feedback was given as to whether the choice was correct. The discrimination task was made harder or easier using a conventional staircase technique, in order to find how much morphing was required to just allow discrimination. If the observer correctly identified the presentation containing the morphed image 5 times, then a new morphed image was chosen which would be harder to discriminate from the reference; if, however, the observer made one or more errors in a sequence of five trials, then an easier morphed image was chosen for subsequent trials. In a single experiment, 4 different spectral slopes were randomly chosen and the thresholds for these were measured concurrently. Two independent staircases were run for each slope.

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

Acknowledgements.

We are very grateful to Dr. P.J. Benson for preparing the man-to-woman morph sequences for us, and to Dr. I.D. Gilchrist for acting as an observer. CAP and the project as a whole were supported by the MRC (Grant Number G9706380)

References

1. H. B. Barlow: **Possible principles underlying the transformation of sensory messages.** In Sensory Communications. Edited by W. A. Rosenblith. Cambridge, Mass.: MIT Press; 1961: 217-234.

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

3. S. B. Laughlin: **Matching coding to scenes to enhance efficiency.** In Physical and biological processing of images : proceedings of an international symposium organized by the Rank Prize Funds, London, England, 27-29 September, 1982. Edited by O. J. Braddick, A. C. Sleigh and Rank Prize Funds. Berlin ; New York: Springer-Verlag; 1983: 42-52.

4. M. V. Srinivasan, S. B. Laughlin and A. Dubs: **Predictive Coding - a fresh view of inhibition in the retina.** Proceedings of the Royal Society of London Series B-Biological Sciences 1982, **216**: 427-459.

5. J. H. Van Hateren: **Theoretical predictions of spatiotemporal receptive-fields of fly LMCs, and experimental validation.** Journal of Comparative Physiology a-Sensory Neural and Behavioral Physiology 1992, **171**: 157-170.

6. J. N. Lythgoe: **Evolution of visual behaviour.** In Evolution of the eye and visual system. Edited by J. Cronly-Dillon and R. Gregory. London: The Macmillan Press Ltd.; 1991: 3-14.

7. M. Weckstrom and S. B. Laughlin: **Visual ecology and voltage-gated ion channels in insect photoreceptors.** Trends in Neurosciences 1995, **18**: 17-21.

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

9. J. J. Atick and A. N. Redlich: **What does the retina know about natural scenes?** Neural Computation 1992, **4**: 196-210.

10. B. A. Olshausen and D. J. Field: **Emergence of simple-cell receptive field properties by learning a sparse code for natural images.** Nature 1996, **381**: 607-609.

11. J. H. Van Hateren and A. Van Der Schaaf: **Independent component filters of natural images compared with simple cells in primary visual cortex.** Proceedings of the Royal Society of London Series B-Biological Sciences 1998, **265**: 359-366.

12. Y. Dan, J. J. Atick and R. C. Reid: **Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory.** Journal of Neuroscience 1996, **16**: 3351-3362.

13. D. J. Tolhurst, S. B. Laughlin and J. S. Lauritzen: **Modelling the variation in the contrast of natural scenes, and the function of contrast normalization in the mammalian visual cortex.** Journal of Physiology-London 1997, **504P**: 123-124P.

14. D. G. Hubel and T. N. Weisel: **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.** Journal of Physiology 1962, 106-154.

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

16. D. J. Tolhurst and I. D. Thompson: **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-Biological Sciences 1981, **213**: 183-199.

17. R. L. DeValois, D. G. Albrecht and L. G. Thorell: **Spatial-frequency selectivity of cells in macaque visual cortex.** Vision Research 1982, **22**: 545-559.

18. G. J. Burton and I. R. Moorhead: **Color and spatial structure in natural scenes.** Applied Optics 1987, **26**: 157-170.

19. D. J. Field: **Relations between the statistics of natural scenes and the response properties of cortical-cells.** Journal of the Optical Society of America a-Optics Image Science and Vision 1987, **4**: 2379-2394.

20. D. J. Tolhurst, Y. Tadmor and T. Chao: **Amplitude spectra of natural images.** Ophthalmic and Physiological Optics 1992, **12**: 229-232.

21. G. E. Legge and J. M. Foley: **Contrast masking in human vision.** Journal of the Optical Society of America 1980, 1456-1471.

22. L. N. Piotrowski and F. W. Campbell: **A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase.** Perception 1982, **11**: 337-346.

23. M. G. A. Thomson and D. H. Foster: **Role of second- and third-order statistics in the discriminability of natural images.** Journal of the Optical Society of America a-Optics Image Science and Vision 1997, **14**: 2081-2090.

24. P. J. Benson: **Morph transformations of the facial image.** Image and Vision Computing 1994, **12**: 691-696.

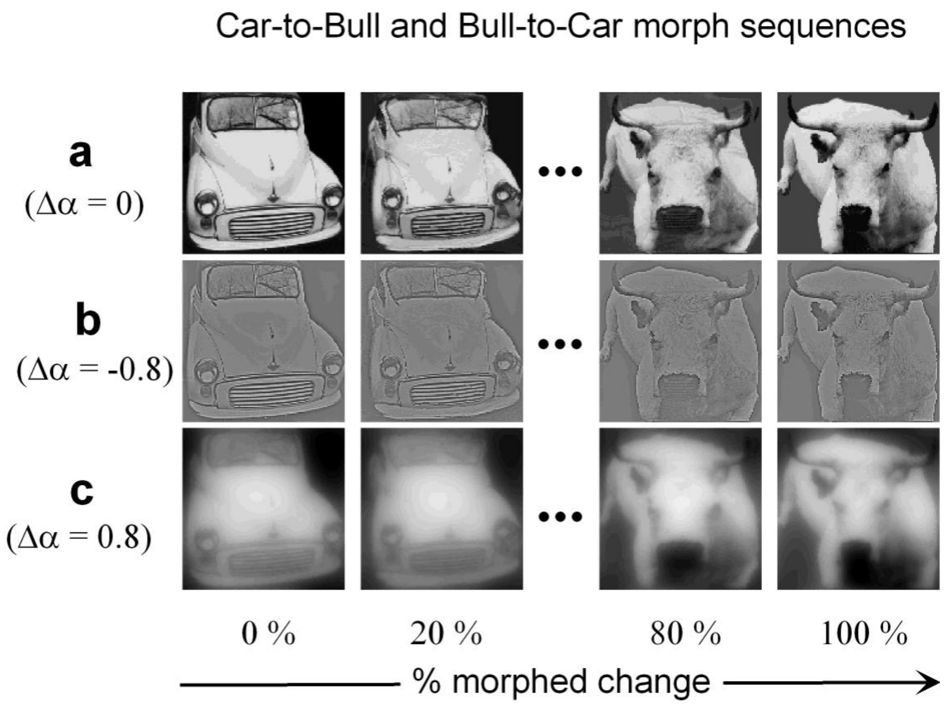
25. D. J. Tolhurst, T. Troscianko, P. J. Benson and C. A. Parraga: **The use of morphed images in the quantitative study of human spatial vision.** Journal of Physiology-London 1998, **506P**: 11-P12.

26. D. G. Pelli and L. Zhang: **Accurate control of contast on microcomputer displays.** Vision Research 1991, **31**: 1337-1350.

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

Figures

Figure 1:



<u>Figure 1:</u> Examples of pictures from two of the four morph series used in our psychophysical experiments. Part **a** illustrates parts from the original car-to-bull morph sequence (left to right) where all pictures have approximately "natural" statistics. The bull-to-car sequence runs right to left. Parts **b** and **c** show the same pictures after they have been processed to decrease ("whiten") and increase (blur) the slopes of their amplitude spectra respectively.

Figure 2:

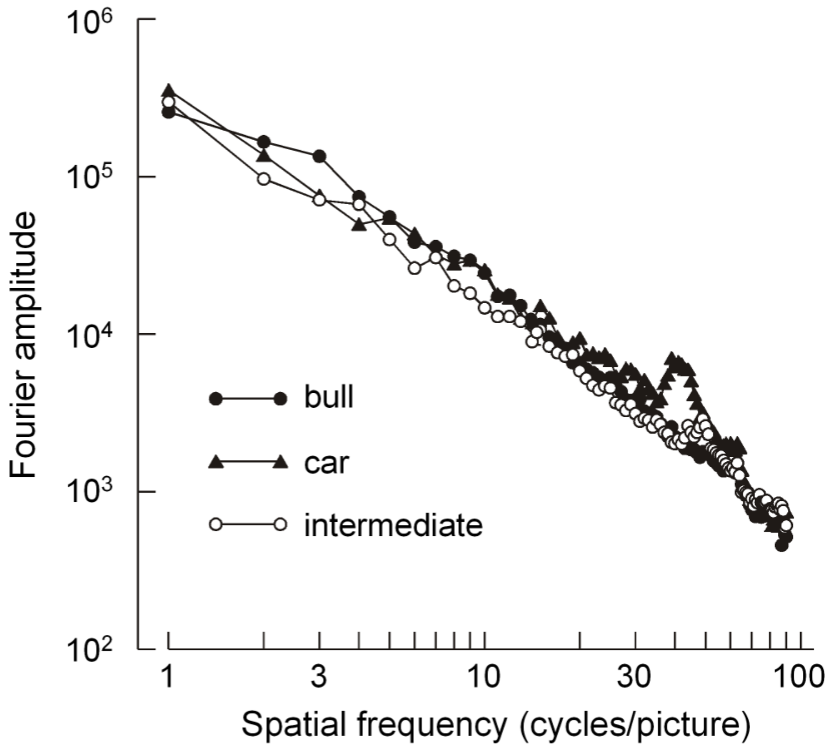


Figure 2: Plots of the spectral amplitude versus spatial frequency for the Fourier spectrum of the Car (0% change), the Bull (100% change) and an intermediate morphed picture from Fig.1a. The plot shows the similarity between the spectra, and hence, the second order statistics of all the pictures. All slopes (α in Eqn.1) fall within a close range (~1.5).

Figure 3:

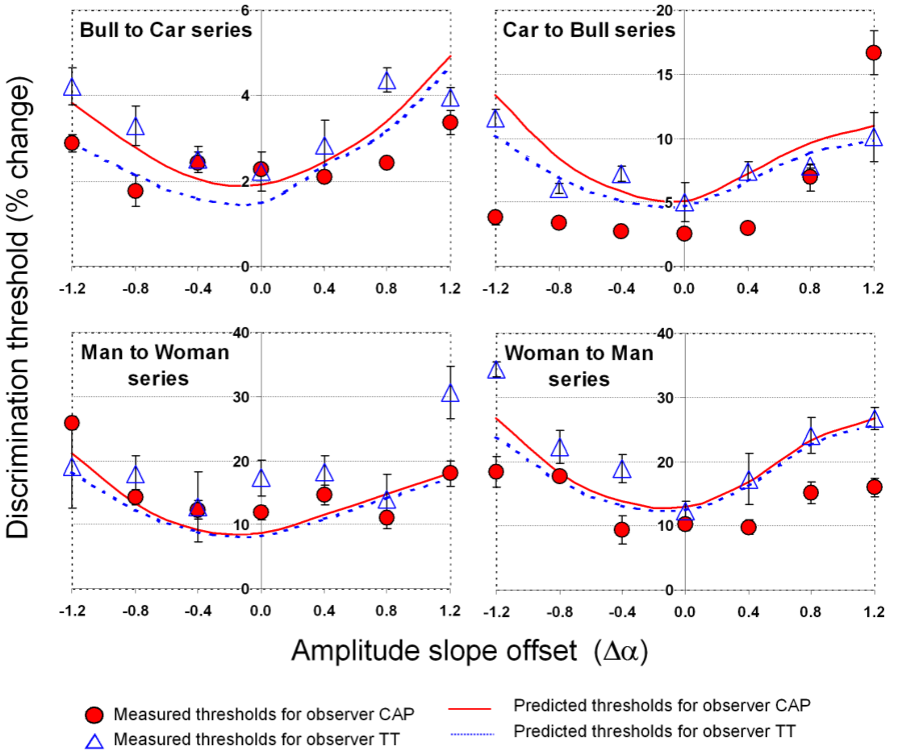


Figure 3: Experimental results obtained for two observers. The plots show the discrimination thresholds for all four morphs sequences versus the deviation ($\Delta\alpha$) from the "natural" value of α. Threshold is expressed as a percentage movement through the morph continuum, with 0% representing one of the two original pictures used to create the morph sequence and 100% representing the other original picture. The error bars show +/- one standard error. The curves show the predictions made by our model of local contrast discrimination.

Figure 4:

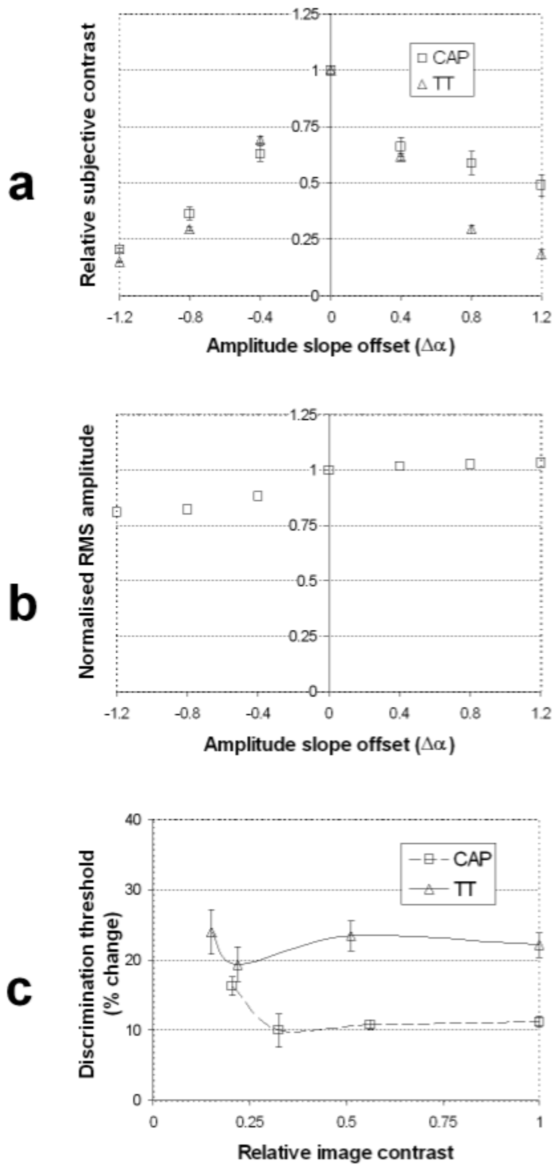


Figure 4: Part **a** shows the relative contrast of pictures of the woman's face when filtered to have different amplitude slope offsets. The observers reduced the contrast of the picture with zero offset (which has unity relative contrast) until it appeared to match the contrast of the pictures with other slope offset. Part **b** shows the normalised RMS amplitude (square root of power) of the same pictures is plotted on the same scale as in **a**; note that subjective contrast (a) is not closely related to physical power. Part **c** shows the results for two observers, who measured their thresholds for discriminating pictures in the woman-to-man series, with pictures presented at a variety of attenuations. The lowest contrasts in this experiment matched the subjective contrasts of picture with slope offsets of +/- 1.2.