

MAGNITUDE RATINGS FOR PERCEIVED CHANGES IN PHOTOGRAPHS OF NATURAL SCENES MAY BE LINEARLY PROPORTIONAL TO DIFFERENCES IN NEURONAL FIRING RATES

David J. Tolhurst¹, Michelle P.S. To¹, P. George Lovell² and Tom Troscianko²
¹*Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, UK* and ²*Department of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK*
Email: djt12@cam.ac.uk

Abstract

We are studying how people perceive suprathreshold changes in the colour, size, shape or location of items in images of natural scenes. We use magnitude estimation ratings to characterise the sizes of the perceived changes in thousands of photographs, and we have built a computational model that tries to explain observers' ratings of naturalistic differences between image pairs. We model the action-potential firing rates of millions of neurons, having linear and non-linear summation behaviour closely modelled on real visual-cortex neurons. Although the model is still imperfect, it does produce tolerable predictions of the ratings for most kinds of image change. Importantly, ratings rise roughly linearly with the model's numerical output, which represents differences in neuronal firing rate in response to the two images under comparison. While rating may not be directly proportional to metrics of stimulus difference, it seems to be proportional to the neuronal response.

Since the earliest electrophysiological recordings of the responses of sensory neurons and the discovery that intensity is coded as action potential frequency (Adrian & Zotterman, 1926), it has been an important question how neuronal response properties relate to human psychophysical performance (e.g. Werner & Mountcastle, 1965; Borg et al., 1966; Barlow & Levick, 1969; Tolhurst et al., 1983; Newsome & Parker, 1998; Chirimuuta & Tolhurst, 2005). The way that neuronal response depends upon stimulus intensity has been of particular interest, since it bears on the pioneering psychophysical ideas of Fechner and Weber. As early as 1931, B. Matthews studied the muscle spindle (a receptor which increases its activity when the muscle is stretched) and he reported: "If the frequency [of action potential firing] be plotted against the logarithm of the load, the points lie very nearly on a straight line.... It has long been held that, as a stimulus increases in geometric progression, the sensation increases in arithmetic progression (Fechner's Law)." That is, he suggested that the logarithmic relation between response and load (i.e. the tension in the muscle's tendon) might underlie Fechner's Law of sensation.

While this looks like a pleasingly straightforward relation between neuronal behaviour and psychophysical performance, the later study of the muscle spindle (P. Matthews, 1972) illustrates important caveats in the search for Laws of sensory coding or perception. First, it turns out that the response of the muscle spindle is determined by muscle *length* and not by *tension*, and it also turns out that tension is logarithmically related to length. Thus, response is actually *linearly* proportional to length, the appropriate measure of stimulus intensity. In trying to deduce relations between neuronal response or psychophysical magnitude estimation and stimulus intensity (Stevens, 1961), it is clearly important to understand what system of measurement is appropriate for the stimulus intensity. Indeed, it may sometimes be the case (as in our present study) that there is no obvious single-dimensional metric to use.

The behaviour of the muscle spindle illustrates a second caveat: of ensuring that the stimuli under study are natural and are in the *natural intensity range*. Muscle spindle response is linearly proportional to muscle length, but only for very small stretches. While this seems, at first, to lessen any interest in a linear response range, it is compatible with the observation that, *in situ*, the length of a muscle can actually change very little (about 5-10%), even though the joints be fully flexed or fully extended. It is important therefore to study sensory systems with stimuli within the natural range.

We have been investigating human threshold and suprathreshold perception of naturalistic changes in digitised photographs of everyday scenes: i.e. natural images (Párraga et al., 2005; Lovell et al., 2006; To et al., 2008): e.g. changes in the shapes, colours or numbers of objects in a scene. We have used magnitude estimation ratings of perceived differences between pairs of images, and we have sought to explain the magnitudes of the ratings with models based on the response properties of neurons in primary visual cortex, V1 (after Watson & Solomon, 1997; Watson & Ahumada, 2005). The great variety of image changes in our experiments means that there is no single physical stimulus metric against which we can compare the observers' ratings; a model of the responses of millions of V1 neurons is an attempt to unify the data. In this paper, we consider what the output of a V1 model tells us about the hypothesised relation between neuronal responses and human magnitude perception.

Methods

Our methods for constructing and presenting stimuli are given in detail by To et al. (2008). 900 pairs of images were made from coloured photographs of natural scenes. Some image pairs could be made by taking one natural photograph and using some kind of image processing technique to change the colour (hue and/or saturation) of all or part of the scene (e.g. Figure 1A). Images could also be blurred or sharpened. Many image pairs were made from a pair of photographs of the same scene. In the time between photographs, the shape or arrangement of objects in the scene may have changed (e.g. Figure 1B), or an object may have appeared or disappeared (e.g. Figure 1C), or the natural lighting and shadowing may have changed. Some image pairs were made by combining the natural shape changes with image-processed colour or blur changes. In fact, in the experiments reported here, we used inverted “negatives” (To et al., 2008) to remove semantic cues, which we have found lead to increased inter-observer variability.

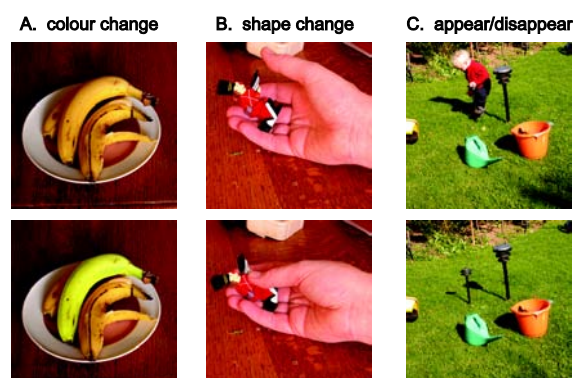


Figure 1. Three examples of the kinds of image pair used in our experiments.

The images in a pair were presented sequentially. The first image was presented for 833ms followed by a 83ms interval when the screen was uniform grey. The second image was then presented for 833ms, followed by a 83ms grey interval and a 833ms re-presentation of

the first image. The observer then gave a numerical rating of the perceived difference between the images. Every 10 trials, one particular image pair was presented (a picture of a red flower where the difference was in colour saturation); the numerical difference between this standard pair was set to “20”, and observers were instructed and trained to use a ratio scale to rate any kind of difference in any other image pair with respect to this standard. Observers could give a rating of “0” if they believed that the images in a pair were identical; there was no upper limit set. At the end of the experiments, each of the 14 observers’ ratings was divided by the median value for the observer, and the normalized ratings for each pair were averaged across observers.

A V1 based model of visual discrimination

Our model is based on that of Watson & Solomon (1997) and Watson & Ahumada (2005) which is used to explain detection thresholds for monochrome grating stimuli. We have tried to extend the models to encompass coloured stimuli (see Lovell et al., 2006) and suprathreshold decisions.

Briefly, the model consists of “simple-cell” receptive fields with Gabor profile – 6 orientations by 5 spatial frequencies by two spatial phase symmetries. Each field is represented at every spatial location in the images, and there are 3 of each field type to deal with image colour (a “luminance” detecting field, and a red-green opponent and a blue-yellow opponent field). The first image in the pair is first convolved with all the different field types, and these linear responses are divided by the local mean luminance values to give *contrast* responses. The r.m.s. is taken of the responses of paired odd-symmetric and even-symmetric “simple cells” to give “complex cell” responses. In the simplest model, the second image would be similarly processed and the responses to the image pair would be compared neuron by neuron. However, realistic models of V1 incorporate threshold behaviour, and non-linear suppressive interactions between neurons (Blakemore & Tobin, 1972; Heeger, 1992). The simple contrast responses of the neurons must be transformed by a sigmoidal transducer function of the form (after Legge & Foley, 1980):-

$$response = \frac{|contrast|^p}{1 + W_N \cdot Normalise^q + W_S \cdot Surround^r} \quad \text{Equation 1}$$

The numerator with power p (about 2-3) gives a positive acceleration, or threshold; the terms in the divisor represent two different kinds of inhibitory interaction between neurons and their effect (Figure 2) is to cause response to become compressed at higher contrasts. The parameters of this equation were adjusted to give the best fit to contrast discrimination experiments (Legge & Foley, 1980) using sinusoidal gratings of various geometries; we then hypothesize that the same sigmoid will describe suprathreshold perceived magnitudes in natural images. Figure 2 shows how the sigmoidal transducer function would lead to something akin to Fechner’s or Weber’s Law for contrast: to obtain a fixed increment in response (DR), greater and greater increments in contrast (DC) are needed as the transducer saturates more at higher contrasts.

The transformed responses of the millions of neurons are compared between the images in the pair, and the millions of differences are pooled by Minkowski summation (Watson & Solomon, 1997). This gives a single number which is the model’s prediction of the observer’s magnitude estimate. We shall discuss below exactly what the y-axis in Figure 2 (the neuronal “response”) actually represents.

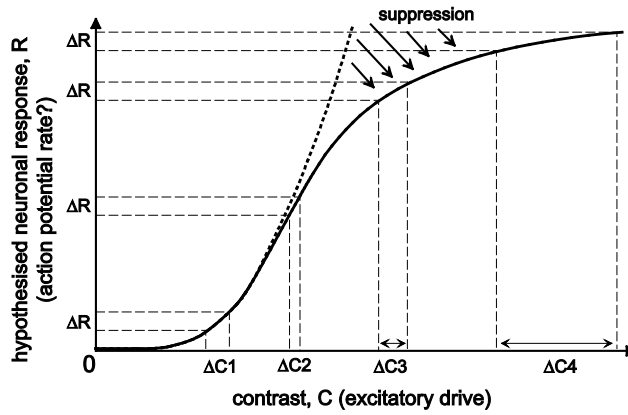


Figure 2. The sigmoidal transducer function in V1-based models of discrimination (solid curve). The dotted curve shows the positively accelerating numerator of Equation 1, before the suppressive interactions between neurons cause the transducer to decelerate.

Results and Discussion

Figure 3A plots the observers' ratings for the 900 inverted and negated image pairs against the predictions of a V1-based model of visual discriminations, a model that relies on the activity of millions of neurons and yet returns only a single number as its output. There is clearly scatter in the fit of the ratings to the model ($r = 0.725$), and there must be more to predicting an observers' ratings than understanding only the low-level coding processes of V1. However, the correlation is adequate enough that we can see that rating is roughly directly proportional to the model's output.

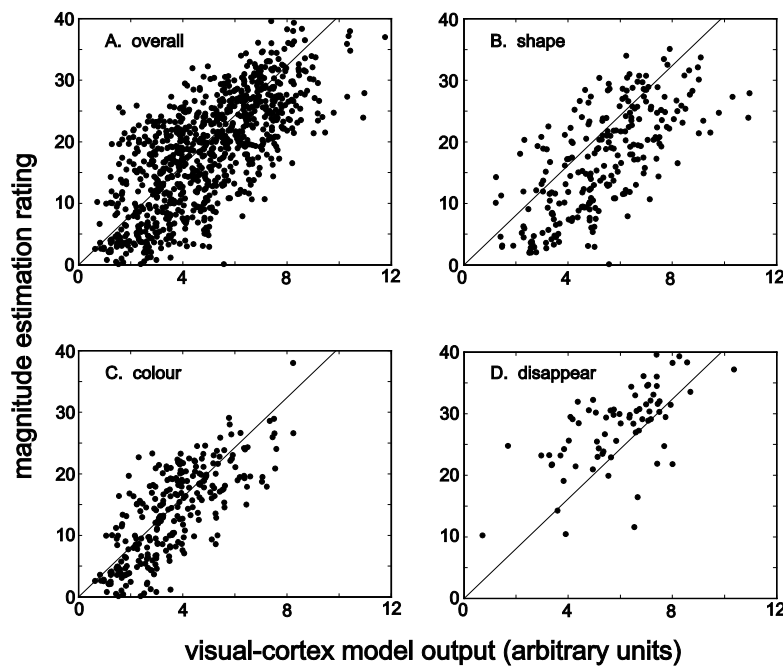


Figure 3: magnitude estimation ratings (average of 14 observers) plotted against the output (in arbitrary units) of a “complex cell” model of visual discrimination. The lines in each panel are the same: the robust regression through the overall dataset.

Figure 3B-D show 3 subsets of the data in Figure 3A, for the classes of image change illustrated in Figure 1. For comparison with the overall trend, the lines are the robust regression through the overall dataset. The data for natural shape changes tend to lie below

the overall trend, while the data for object disappearance tend to lie above; perhaps, the trend for colour changes (Figure 3C) is steeper than the overall trend. These differences in the relations for different image change types contribute to the scatter in the overall (Figure 3A), and they show that we still have not succeeded in providing a model that completely explains the magnitude estimation ratings given for naturalistic changes in natural images.

It has been argued whether magnitude estimation ratings need be directly proportional to the internal magnitude of sensation (e.g. Gescheider, 1997), but the present results suggest that the ratings depend linearly on neuronal response levels. The numerical output of our V1 modelling (Watson & Solomon, 1997) reflects differences in the magnitudes of neuronal responses to the two image under comparison. At first sight, it would seem that the perceived magnitude difference between two natural images is directly proportional to the difference in neuronal response to the two images, where response might be expressed as total number of action potentials generated during an image presentation. Boynton et al. (1999) have argued that the sigmoidal transducer function which is presumed to underlie contrast discrimination thresholds is the same shape as the relation between the V1 BOLD (fMRI) signal and contrast; while Heeger et al. (2000), in turn, argue that the BOLD signal follows the relation between neuronal action potential rate and contrast.

However, the hypothesized sigmoidal transducer function (like Equation 1; Figure 2) does not simply describe the relation between response amplitude and contrast for single V1 neurons. Individual V1 neurons each respond to very limited ranges of contrast, while the dynamic ranges of different neurons cover different contrast ranges (Tolhurst et al., 1981, 1983; Albrecht & Hamilton, 1982; Sclar et al., 1990; Heeger et al., 2000). Thus, the transducer is the pooled response of many neurons and its shape reflects how many neurons are responsive at each contrast rather than the shape of the response-contrast functions of single neurons (Clatworthy et al., 2003; Chirimuuta & Tolhurst, 2005).

It is well accepted that human detection thresholds or intensity discrimination limens involve statistical judgments about changes in an “internal variable” (Green & Swets, 1966), and it is a common observation in sensory neurophysiology that neuronal response variability (or “noise”) increases with response magnitude and presumably with stimulus intensity (Werner & Mountcastle, 1965; Tolhurst et al., 1981, 1983, 2009). The “response” metric in the sigmoidal transducer is not action potential rate *per se*, but must incorporate the increasing neuronal “noise” at higher contrasts.

Acknowledgements

This research was supported by grants from the EPSRC and Dstl on the Joint Grants Scheme.

References

- Adrian, E.D. & Zotterman, Y. (1926). The impulses produced by sensory nerve endings. *Journal of Physiology*, 61, 151-171.
- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, 48(1), 217-237.
- Barlow, H.B. & Levick, W.R. (1969). Three factors limiting the reliable detection of light by retinal ganglion cells of the cat. *Journal of Physiology*, 200, 1-24.
- Blakemore, C., & Tobin, A. (1972) Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, 15, 439-440.
- Borg, G., Diamant, H., Strom, L. & Zotterman, Y. (1966). The relation between neural and perceptual intensity: a comparative study on the neural and psychophysical response to taste stimuli. *Journal of Physiology*, 192, 13-20.

- Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neural basis of contrast discrimination, *Vision Research*, 39, 257-269.
- Chirimuuta, M. and Tolhurst, D.J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research*, 45, 2943-2959.
- 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.
- Gescheider, G. A. (1997). *Psychophysics – The Fundamentals*. USA: Lawrence Erlbaum Associates.
- Green, D.M. & Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. Wiley.
- Heeger, D. J. (1992) Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181-197.
- Heeger, D. J., Huk, A. C., Geisler, W. S., & Albrecht, D. G. (2000). Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nature Neuroscience*, 3, 631–633.
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70, 1456-1471.
- Lovell, P.G., Párraga, C.A., Ripamonti, C., Troscianko, T. and Tolhurst, D.J. (2006). Evaluation of a multi-scale color model for visual difference prediction. *ACM Transactions on Applied Perception*, 3, 155-178.
- Matthews, B.H.C. (1931). The response of a single end organ. *Journal of Physiology*, 71, 64-110.
- Matthews, P.B.C. (1972). *Mammalian Muscle Receptors and their Central Actions*. Hodder & Stoughton.
- Parker, A.J. & Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Reviews of Neuroscience*, 21, 227-277.
- Párraga, C.A., Troscianko, T. & Tolhurst, D.J. (2005). The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model. *Vision Research*, 45, 3145-3168.
- Sclar, G., Maunsell, J.H. & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30, 1-10.
- Stevens, J.J. (1961). To honor Fechner and repeal his Law. *Science*, 133, 80-86.
- To, M., Lovell, P.G., Troscianko, T. & Tolhurst, D.J. (2008). Summation of perceptual cues in natural visual scenes. *Proceedings of the Royal Society B*, 275, 2299-2308.
- Tolhurst, D.J., Movshon, J.A. & Thompson, I.D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, 41, 414-419.
- Tolhurst, D.J., Movshon, J.A. & Dean, A.F. (1983). The statistical reliability of signals in single neurones in cat and monkey visual cortex. *Vision Research*, 23, 775-785.
- Tolhurst, D.J., Smyth, D. & Thompson, I.D. (2009). The sparseness of neuronal responses in ferret primary visual cortex. *Journal of Neuroscience*, 29, 2355-2370.
- Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision* 5(9), 717-740, doi: 10.1167/5.9.6.
- 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.
- Werner, G. & Mountcastle, V.B. (1965). Neural activity in mechanoreceptive cutaneous afferents: stimulus-response relations, Weber functions, and information transmission. *Journal of Neurophysiology*, 28, 359-397.