A Spatiochromatic Model of Vision

Ian R Moorhead
Vision & Visual Systems Section, WXI Depastment, Defence Research Agency
Fort Halstead, Sevenoaks, Kent, United Kingdom, TN14 7BP

Abstract

A computer model of human spatiochromatic vision, based on the scheme proposed by De Valois and De Valois[1] has been developed. The implementation of the model enables true colour 2-D images to be processed. The input consists of cone signals at each pixel. Subsequent levels of the model are represented by arrays of activity corresponding to the equivalent neural activity. The implementation allows the behaviour of different stages of the model—retinal and cortical—to be studied with different varieties of spatial and chromatic stimuli of any complexity. In addition the model is extensible to allow different types of neural mechanisms and cortical demultiplexing processes[2] to be incorporated.

As well as providing qualitative insight into the operation of the different stages of the model the implementation also permits quantitative predictions to be made. Both increment threshold and hue naming results are predicted by the model, but the accuracy of these predictions is contingent upon an appropriate choice of adaptation state at the retinal cone and ganglion cell level.

Keywords: Colour Vision, Increment Threshold, Spatiochromatic Filtering, Demultiplexing

2. Introduction

The research described in this paper forms a component of a programme of work concerned with developing a system for predicting human visual performance. The aim is to devise what I term a "second-generation" visual performance model. Typically, current visual performance models are parametric (eg Guth,[3] Blackwell[4]) requiring the user to abstract appropriate descriptors of the scene or image for which the estimate of performance is desired. The aim of the present work is to be able to produce a model which accepts as input, not parameters, but the image itself and hence generate an estimate of performance for the desired task. In order to simplify matters somewhat, colour processing has been separated from luminance processing initially (although the final model will integrate these and other modules). This paper describes preliminary work with the colour model.

The visual world varies in both colour and luminance and many visual systems have evolved to exploit these variations to obtain information about objects in a scene. Although colour and luminance covary in an image (eg Burton and Moorhead[5]) the majority of research has tended to concentrate on either the processing of spatial (ie luminance) signals, or the processing of colour signals. The reasons for this were pragmatism—by separating out colour and luminance it was easier to understand the visual mechanisms involved. Also the trichromatic theory of colour vision was extremely successful without recourse to modelling spatial processes.

However, some attempts have been made to understand how colour and spatial information is jointly processed by the early stages of vision. Early work by van der Horst[6] measured the contrast sensitivity of the different colour channels. More recently Losada and Mullen[7] have measured the characteristics of chromatic channels. Atick et al[8] using an information theoretic approach to visual processing have shown that receptive fields that are spatially and spectrally selective are optimal for processing natural scenes. Ingling and Martenez-Uriegas[9] showed theoretically that simple opponent receptive fields could be viewed as multiplexing together luminance and colour information, in such a way that colour information was effectively handled as a lowpass spatial signal, while luminance information was carried as a bandpass signal. Since psychophysical experiments[10] have shown that human vision appears to have different characteristics when presented with purely colour information, compared with purely luminance information, it has become fashionable to consider that a demultiplexing process operates at cortical level to separate out colour and luminance information from the outputs of the retinal ganglion stage.

Recently De Valois and De Valois[1] published a model that suggests an explicit three stage model of how the retina combines information from the different cone types at the retina and how cortical units then extract separate luminance and colour channels from this representation. The model is linear apart from a final rectifying process and brings together ideas from previous work (Lennie,[11] D'Zmura & Lennie[12]). The model differs from previous ones in that the authors propose that linear combinations of the common "red-green" single opponent operators are modulated by a single opponent operator receiving inputs from short wavelength cones. Depending upon which linear combinations are taken then determines the spatial and spectral characteristics of the derived channels.

The work reported here was a preliminary attempt to produce a computer simulation of the De Valois and De Valois model in a form that is able to directly process full 2-D colour images. It is necessarily simple at present and does not include any form of chromatic aberration. The purpose was to examine firstly whether the mechanism for demultiplexing colour and luminance information would work as suggested, and secondly to determine whether the signals from the colour channels might be useful for generating colour descriptors for natural scenes. Even this simple implementation of the model highlighted the fact that a number of aspects of the model had not been made explicit in the original paper.

Section 3 summarises the De Valois and De Valois model briefly and explains a number of aspects that needed to be quantified in a computational version. Section 4 illustrates how the different stages of the model respond qualitatively to a simple colour image. Section 5 shows the results of simulating an increment threshold experiment and illustrates that the model can be used to predict threshold performance.

3. The Model

3.1 Description

This section will summarise the main components of the De Valois and De Valois model and highlight certain modifications and additions that were necessary to create a working computational scheme. The principle components of the model are illustrated schematically in figure 1. The input is a retinal image consisting of signals from the three types of cone. The relative proportions of the long (L), medium (M) and short (S) cones is specified to be 10:5:2. In the retina it is only physically possible to have a single cone at any one location, but for simplicity at present in the computational version each cone type assumed to existed at every location. The inputs to the next stage were weighted according to the relative abundance of each cone type. (It can be argued that the point spread function of the eye effectively spreads the same signal over a number of cones[13]). Future work will examine the effect of only having a single cone type at each location in the input. Cone spectral sensitivities corresponded to the Smith and Pokorny fundamentals.[14]

The second stage of the model corresponds to the retinal ganglion cells. Receptive fields are constructed which have a single cone type input to the centre and one or more cone types as inputs to the surround. De Valois and De Valois designated these units according of the type of cone that formed the input to the centre of the receptive field, thus a unit with L cone inputs to the centre is an L_o unit. Similarly M_o and S_o units receive inputs to the centre of their receptive fields from M cones and S cones respectively. The centre and surround are antagonistic. Figure 2 illustrates schematically the kinds of units that result. The spatial structure of the receptive field was arranged to be a difference of Gaussian, to conform with other models of ganglion cells.[15] The units at this stage are of two polarities—ON units with receptive field centres which are excited by a particular cone type and OFF units whose centres are inhibited by a particular cone type. When actually producing a computer implementation of spatiochromatic operators of this kind it is necessary to decide exactly how the centre and surround responses will be weighted relative to each other. In the case of models of spatial vision, difference of Gaussian operators are generally arranged to be balanced ie to produce no response to a uniform field (eg Watt[16]). This concept can be generalised to spatiochromatic operators like those described here, and they can be set up to produce a null response to a uniform field *of a particular colour*. If R_c represents the response from the centre of the receptive field and R_s the response from the surround, then we require that $R_c = R_s$ to achieve this. These can be computed by summing spatially and spectrally, for each cone contribution to that part of the receptive field, thus:

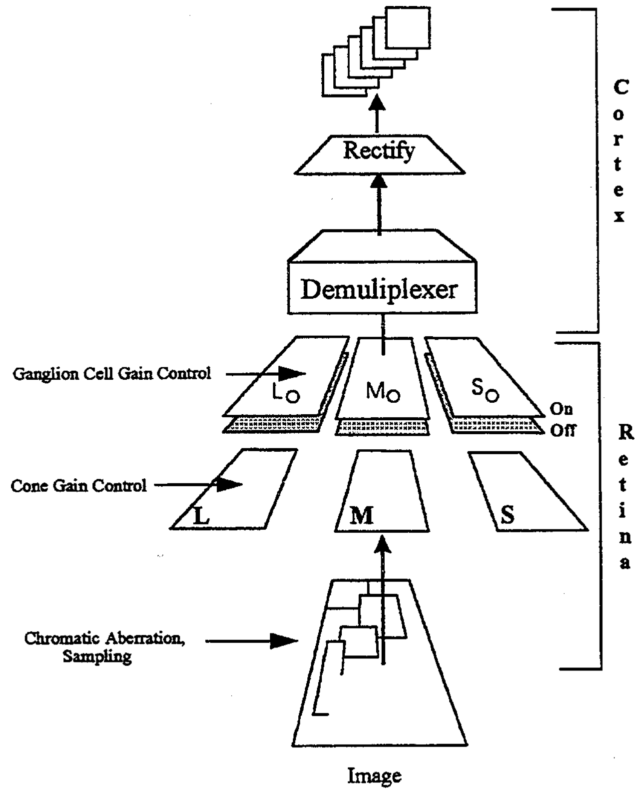

Figure 1. Diagram illustrating the major components of the model.

The diagram in Figure 4 illustrates the components of the model. There are 4 main stages:

1) Photoreceptors; L, M and S (Smith and Pokorny)
2) Single opponent operators; L_o, M_o and S_o which can be either 'ON' or 'OFF' units
3) Demultiplexing by linear superposition
4) Rectification

Each component can be individually modified. Also shown are factors that are important to the operation of the model and which were not incorporated in the original formulation of DeValois and DeValois. These are:

1) Chromatic aberration
2) Cone Gain Control (local)
3) Ganglion Cell Gain Control

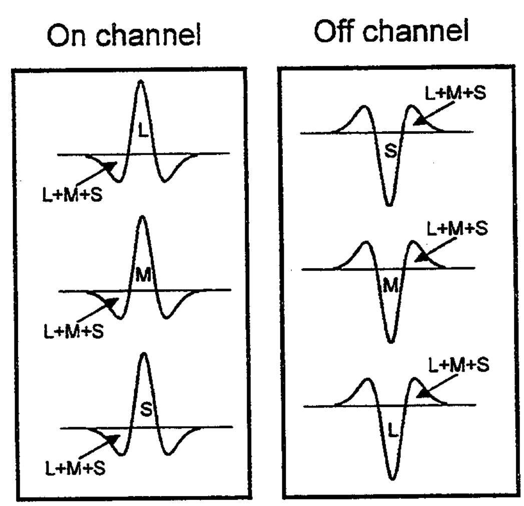

Figure 2. Schematic illustration of the single opponent processes corresponding to retinal ganglion cells. In the initial formulation only a single cone type contributed to the centre response, whereas all three cones had inputs to the surround.

$$R_c \sum_{x \in centre} \sum_{i=1}^{3} \int_{380}^{750} c_i(x,\lambda)S(x,\lambda)d\lambda$$

$$R_s \sum_{x \in surround} \sum_{i=1}^{3} \int_{380}^{750} c_i(x,\lambda)S(x,\lambda)d\lambda$$

where $c_i(\mathbf{x}\ \lambda)$ is the spectral sensitivity of the ith cone at position $\mathbf{x}$ and wavelength λ and $S(\mathbf{x}, \lambda)$ is the spectra power at position $\mathbf{x}$ and wavelength λ of the stimulus that is being used to weight the centre and surround. Note that this general formulation makes no assumption about which types of cone contribute to the centre and surround. Whilst computationally this scheme is easy to achieve, it is not so clear that retinal ganglion cells operate in the same manner. The process in fact is a simple form of local colour constancy. This aspect of the model was not considered by De Valois and De Valois.

The next stage of the model simulates the process of demultiplexing the colour and luminance information encoded by the retinal ganglion cell signals. Figure 3, adapted from De Valois and De Valois sunrises the process. What happens is that linear combinations of the various types of single opponent units are used to generate two lightness channels and four colour channels. The basic process combines L_o and M_o units. These are then further modulated by contributions from S_o channels. The linear combination stage is followed by a non-linear half-wave rectification process. The linear combination process proposed by De Valois and De Valois requires that the single opponent L_o, M_o and S_o units be combined in a particular ratio (10:5:2). If this is not done the model no longer provides an accurate prediction of the unique hue values. It is not made explicit in their model whether these units are combined from a single location in the visual field or whether, for example ten L_o units are summoned over a region of the image. Again for simplicity the units were combined at a single spatial location, but with the appropriate weighting factor.

3.2 Implementation Details

The input to the computational version consisted of three arrays (which could be of any size) of cone signals. These were computed using image spectral information (CIE tristimulus values were computed and then transformed to the corresponding Smith and Pokorny cone signals in the conventional manner). Generally images consisting of 64×64 or 128×128 values were used to reduce the time of computation. The single opponent filters were 15×15 pixels in size with a standard deviation for the centre of 0.68 pixels (this corresponds to a 1.5 octave (HWHH) filter with a centre frequency of 0.32 cycles/pixel). The standard deviation of the surround was twice this value. Responses were computed by direct convolution from the cone "image" arrays.

4. Qualitative Behaviour

In this section I will demonstrate qualitatively how the different stages of the model respond to a simple coloured image. The image consisted of a set of coloured tiles, arranged as a grid consisting of six rows and four columns. For the 128×128 pixel image used as an example each tile was 32×21 pixels (this leaves a small margin at the top and bottom of the image). The reflectance specimen of each tile was chosen from the set of Munsell colours. The tiles were assumed to be uniformly illuminated by a D_{65} illuminant of a particular luminance. For the data that will be presented here the image consisted of tiles which all had the same red hue (R2.5, in Munsell notation), but varied in value and chroma. The tiles making up the image were arranged so that the same set of 12 colours were present in the top half of the image as the bottom half and had the same arrange-

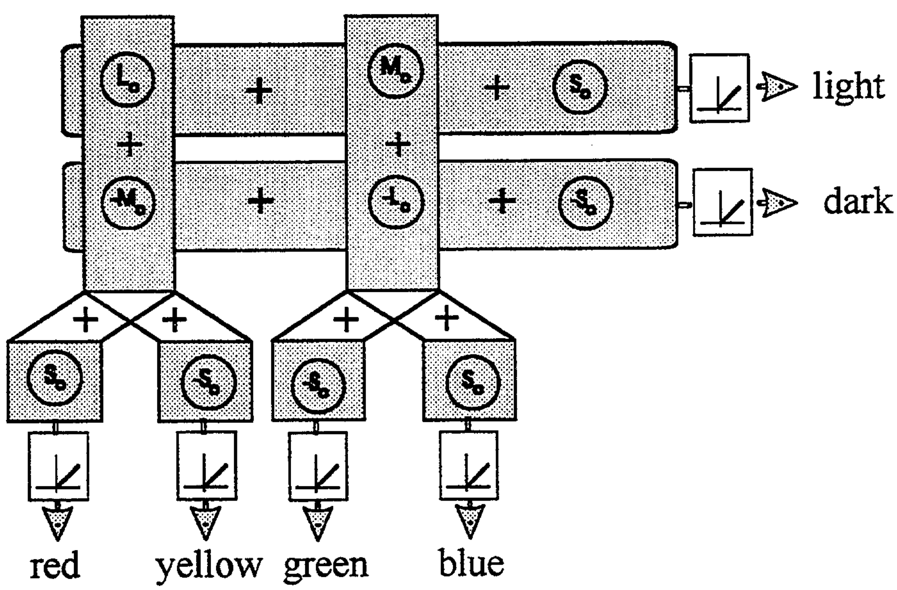

Figure 3. Demultiplexing scheme proposed by De Valois and De Valois. Summation of like sign single opponent units creates lightness and darkness responses, while summation of opposite sign units creates colour channels.

ment. This was used in order to allow the top half of the image to be set to be isoluminant (for the CIE standard observer). This image structure allows one to examine the behaviour of the model under normal and isoluminant conditions simultaneously.

The responses of the different stages of the model are illustrated in figures 4 and 5. The plot on the left of figure 4 shows the response of the L cones to this simple test image. The front (X) axis corresponds to the bottom of the image. One can see that the cone signals directly reproduce the tiled structure of the input image. This is because the model does not at present incorporate simulation of either the optical point spread function or of the sampling by the different receptor types. The three rows of tiles at the bottom of the image which were set to be isoluminant, produce very similar responses from the fact that the cones are responding to colour differences only, whereas the top three rows are responding to colour and luminance differences jointly.

The plot on the right of figure 4 illustrates the responses of a single opponent L_o channel to the same test image. In the region of the graph corresponding to the upper part of the image the responses are typical of a centre-surround single opponent filter. There is a mixture of band-pass, "edge" responses, resembling those produced by a conventional luminance difference of Gaussian filter, and low-pass "DC" responses. These operators produce particularly strong responses at corners when there are both luminance and colour differences as indicated by the large spikes. In the isoluminant portion of the image where there are now no luminance edges only the "DC" responses can be seen. This graph re-emphasises the fact that these kinds of operators multiplex together low-pass colour signals and band-pass luminance signals. The responses of the other types of single opponent operators are similar. It is the function of the next stage to separate out these two types of response.

If the subsequent demultiplexing stage operated to separate these two signals the response of the colour channels should appear as a low-pass representation of the input image. Figure 5 on the left shows the response from the "lightness" channel. In the region where the image contained luminance information the responses are close to what one would expect from a simple luminance operator. However, over all of the image there are small DC responses, in particular in the isoluminant portion of the image. Although small it appears that the De Valois and De Valois "lightness" channel also has a response even when the stimulus is isoluminant Similarly, the response of the "red" channel illustrated in the plot on the right of figure 5 shows considerable "bandpass" operator like responses. It is clear that there are still significant "edge transients" present in the bottom part of the image which contains both luminance and colour information. Not surprisingly the isoluminant part of the image does show a low pass representation of the input image.

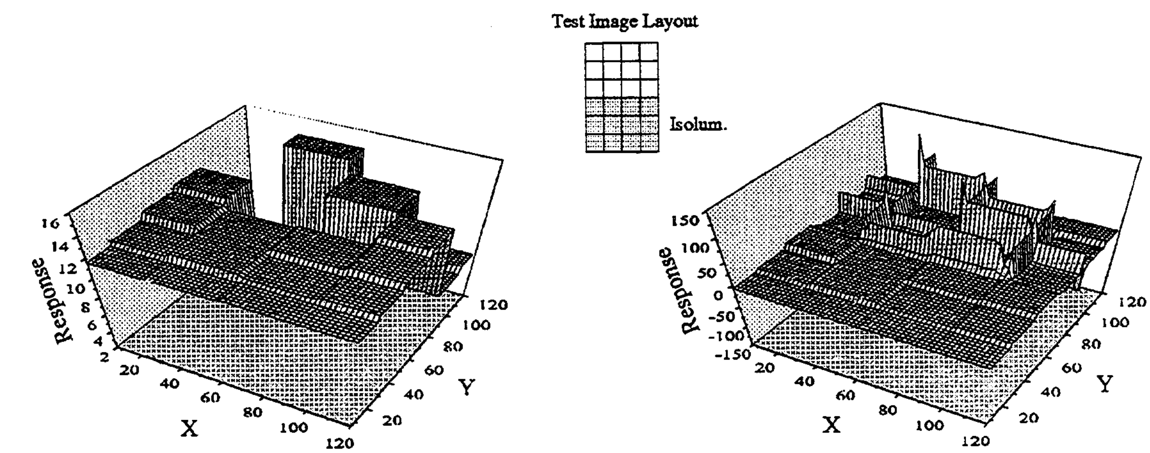

Figure 4. Surface plots of (a) the responses of the L cones and (b) the L cone centre single opponent units. The front margin of the plots corresponds to the bottom of the test image. The small schematic in the centre of the figure illustrates the layout of the test image, consisting of a 4 × 3 grid of Munsell chips, replicated in the lower half of the image and set to be isoluminant.

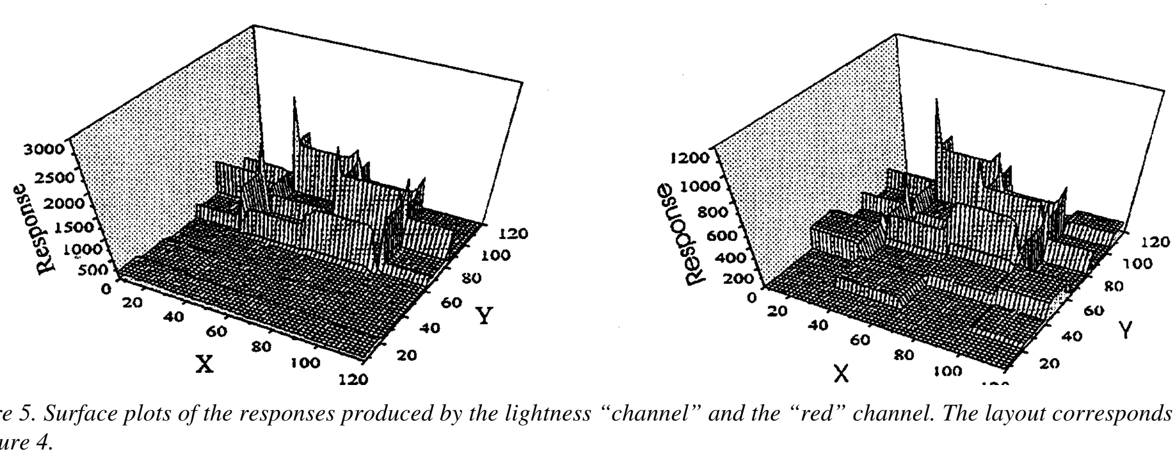

Figure 5. Surface plots of the responses produced by the lightness "channel" and the "red" channel. The layout corresponds to that of figure 4.

These results demonstrate that the demultiplexing process is incomplete. By experimenting with the relative weighting of the single opponent channel inputs to the demultiplexing stage of the model it was possible to achieve much better separation of the colour and luminance signals. To achieve this equal contributions from the L_o and M_o units, without contributions from the S_o units were used. Unfortunately, this new scheme no longer provided an accurate prediction of unique hue wavelengths which is one of the claims of the De Valois and De Valois formulation.

5. Quantitative Example

In order to examine whether the model could reliably predict actual performance an increment threshold experiment was simulated. This section describes the methods and results.

Increment threshold experiments generally involve an observer detecting a monochromatic increment on a white background. This was simulated as follows. A 64×64 pixel uniform white image was created, with a luminance of 50Cd/m^2. The colour temperature of the image could be set to any value required. A circular test spot measuring 10 pixels in diameter was added to this adapting field. The test spot could be of any wavelength and had a Gaussian spectral distribution with a standard deviation of 2nm. The radiance of the test spot could be set also, and for the data presented here it was set arbitrarily to $1.0 \times e^6$ units. Images were created for test spots with centre wavelengths ranging from 400 to 700nm in steps of 10nm. The final output from the model as so far described is a set of 6 arrays corresponding to the colour and lightness channels. In order to obtain a single number that could be used to estimate the response of the model to the test stimulus a number of different schemes were implemented. The simplest involved taking the maximum value from all of the channel responses at the centre of the test spot. A second method computed a root mean square value of all of the channel outputs at the centre of the test spot. Finally a Gaussian weighted spatial average of the root mean square signal was also computed. This could either include just the colour channels or both the lightness and colour channels. The standard deviation of the Gaussian was set to be 2.0 pixels.

Only a single test image was created and the response computed. No modelling of particular psychophysical techniques such as the method of constant stimuli or a two alternative forced choice procedure was included.

Figure 6 illustrates the output from the "green" channel to this kind of test stimulus. On the left is the response to a test wavelength of 470 nm, while on the right is the response to a test wavelength of 570 nm. The responses are quite different in character with that to 570nm being much more like an edge response than that to 470 nm.

The performance results of one such experiment are plotted in figure 7. The background colour temperature in the test images was set to 6500K (D_{65}). The figure on the left shows the responses computed using the different criteria described, as well as a typical set of human psychophysical data obtained by Kranda and King-Smith.[17] All of the data

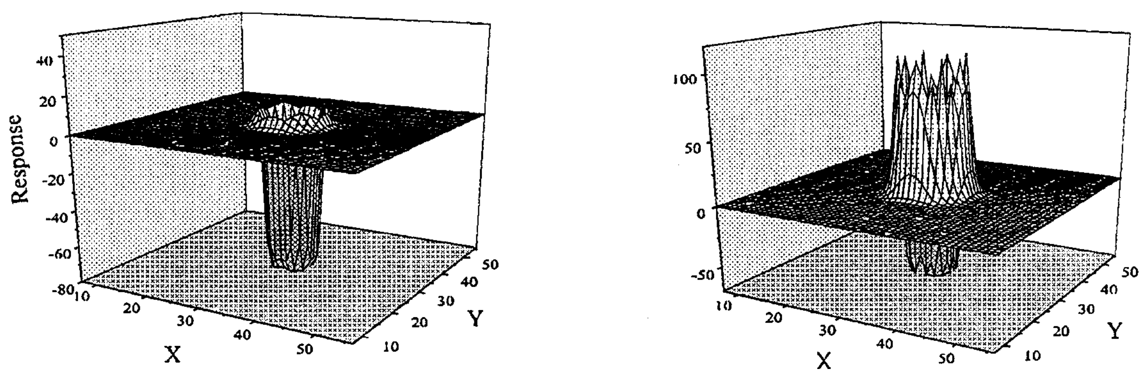

Figure 6. Responses from the "green" channel to monochromatic increment test spots. (a) 470nm test. (b) 570 nm test.

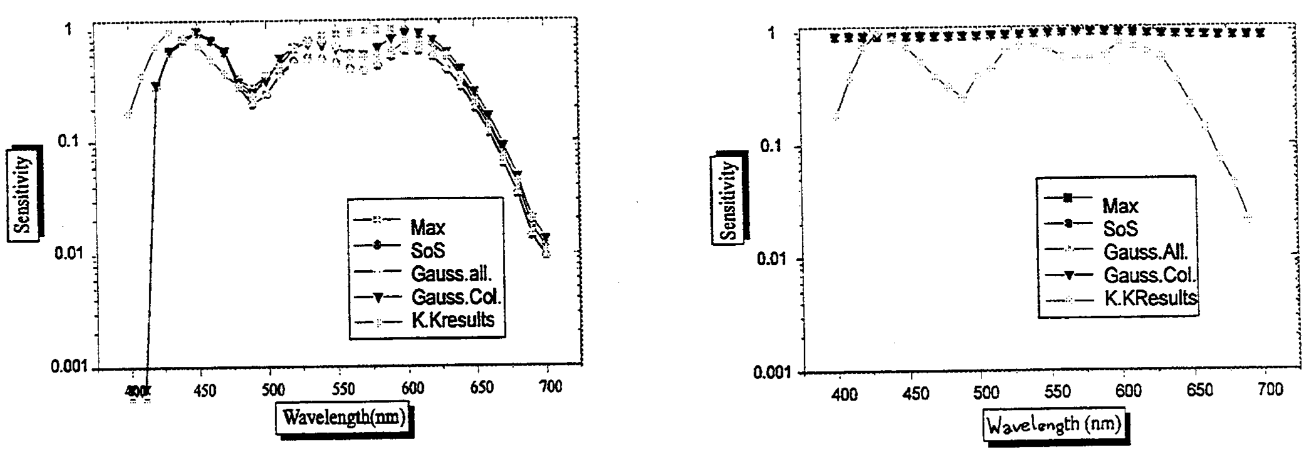

Figure 7. Computed increment thresholds compared with psychophysical measurements. (A) Results obtained when the a single opponent channel receptive fields were normalised to produce a null response to the background. (B) Results obtained when these units were not normalised.

have been normalised before plotting. The data from the model reproduce the three-peaked increment threshold response typically found with human observers. In particular the responses calculated as a Gaussian spatially weighted root mean square of all the channels match the psychophysical data closely including the reduced sensitivity referred to as Sloan's Notch, although the fit is less satisfactory towards the short wavelength end of the spectrum. In fact all of the model responses show a poor fit below wavelengths of 470nm. In addition the method of computing the maximum response does not reproduce the dip (Sloan's notch) in the psychophysical data at around 570nm.

These results are only achieved if the single opponent channel responses are normalised in the manner described in section 2. The plot on the right of figure 7 shows the predictions made by the model for the same increment threshold experiment when the normalisation process is not used. Here the responses are virtually independent of wavelength.

A number of experiments were carried out in which the various parameters were manipulated. In general the results appear to be quite robust as long as the single opponent channels are balanced in the manner described earlier. For example if the colour temperature of the adapting field is set to that of illuminant A, the relative heights of the peaks at 550nm and 620 nm are reduced compared with the blue peak. This is consistent with the fact that illuminant A has a lower colour temperature than D_{65} and therefore produces more adaptation at the red end of the spectrum.

6. Discussion and Conclusions

The model described here is tantalising. In some respects it appears to represent well the processes of human colour vision, and yet on the other hand it makes erroneous predictions. As already shown by De Valois and De Valois the model predicts the unique hue locations correctly. Here I have shown that it can also provide a close approximation to the increment threshold function. On the other hand the predictions made in the hue naming experiment are less convincing. In addition the qualitative results of processing a simple tiled image suggest that the demultiplexing process is incomplete. Finally the quantitative data appear to rely upon the correct balancing of the centre and surround of the single opponent responses in order to provide good predictions of performance. This aspect was not apparent in the original model. What has not been looked at yet is the whether the model can predict effects of changing the spatial parameters of the stimulus, although some preliminary work examining increment threshold performance for small test spots suggests that this formulation of the model does not predict the shift to a V_1 shaped function as occurs psychophysically.

There are a number of parameters that can be manipulated in this model. The data presented here represent a first attempt at setting them. It seems likely that attempts to model other sets of psychophysical data will more strongly define the possible ranges of these parameters.

The incomplete demultiplexing of colour and luminance is problematic. It may be that there is a more appropriate linear combination of the single opponent responses that would not only separate the colour and luminance responses, but would leave the unique hue predictions intact. Alternatively it may be that a non-linear mechanism is required to achieve correct demultiplexing. A third possibility however is that the model actually represents how human vision works. In this case there may be stimuli and experiments that support the idea of incomplete demultiplexing of colour and luminance signals.

Although the prediction made for the increment threshold experiments is very good over the middle to long range of the spectrum, the prediction is poor at the short wavelength end. Not only is the blue peak in the wrong place, but the sensitivity is significantly poorer than that measured psychophysically. It is possible that these discrepancies arise from the fact that the S cone single opponent unit is an inappropriate model. There is some evidence that only certain types of S cone unit exists, and these may well have different sensitivities compared to the L cone and M cone units. Kranda and King-Smith[17] for example required a different slope for the blue-yellow channel when fitting an opponent channel model to their data. Future work will examine some of the uncertainties in the present version of the model and will extend the quantitative predictions to experiments in which space and colour effects interact.

7. References

1. R L. De Valois and K. K. De Valois, "A multi-stage color model", *Vision Res.* **33**:1053 - 1065, (1993).
2. V. Billock, "Cortical simple cells can extract achromatic information from the multiplexed chromatic and achromatic signals in the parvocellular pathway", *Vision Res.* **35**:2359 - 2369 (1995).
3. S. L. Guth, " Model for color vision and light adaptation", *J.O.S.A. (A)* **8**:976-993 (1991).
4. H. R. Blackwell "Description of a comprehensive family of relative contrast sensitivity (RCS) functions of luminance to reflect differences in size of task detail, task eccentricity and observer age", *J. Illum. Eng. Soc.* **11**: 52-63 (1981).
5. G. J. Burton and I. R. Moorhead, "Color and spatial structure in natural scenes", *Applied Optics* **26:** 157 - 170 (1987).
6. G. J. C. Van der Horst, "Transfer of spatial chromatic contrast at threshold in the human eye", *JOSA* **59**:1260-1266 (1969).
7. M. A. Losada and K. T. Mullen, "The spatial tuning of chromatic mechanisms identified by simultaneous masking", *Vision Res.* **34**: 331 - 341, (1994).
8. J. J. Atick, Z. Li and A. N. Redlich, Understanding retinal color coding from first principles", *Neural Comp.* **4**:559 - 572 (1992).
9. C. R. Ingling Jr. and E. Martinez-Uriegas, "The spatiotemporal properties of the r-g X-cell channel", *Vision Res.* **25**: 33-38 (1985).
10. P. Gouras, "Color Vision", *Prog. in Retinal Res* **3**:227-261, (1984).
11. P. Lennie and M. D'Zmura, 'Mechanisms of Color vision", *CRC Critical Reviews in Neurobiology* **3**: 333-400 (1988).
12. M. D'Zmura and P. Lennie, "Mechanisms of color constancy", *J. O. S. A. A* **3**:1662 - 1672, (1986).
13. N. D. Haig, "Local gain control and focal accommodation in the self similar stack vision model", *Spatial. Vision* **7**:15-34 (1993).
14. V. C. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500mn", *Vis. Res.* **15**:161-171 (1975).
15. R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli", *Vision Res.* **5**:583-601 (1965).
16. R. J. Watt, *Visual Processing: computational, psychophysical and cognitive research,* Lawrence Erlbaum Associates, 1988.
17. K. K. Kranda and P. E. King-Smith, "Detection of coloured stimuli by independent linear systems", *Vision Res.* **19**:733 745 (1979)

published previously in SPIE, Vol. 2657, page 10