COLOUR PRINCIPLES AND EXPERIENCE FOR COMPUTER GRAPHICS

Susan M. Goetz, John C. Beatty and Deryl J. Rasquinha

Computer Graphics Laboratory
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
(519) 886-1351

ABSTRACT

The increased availability of colour in raster graphics displays argues a need for greater understanding within the computer graphics community of the physiology and psychophysics of colour. An understanding of the fundamentals governing colour perception provides insight into certain visual colour phenomena which occur, and can provide a basis for colour selection and use. A program for experimenting with colour is an effective means for vividly experiencing these effects first hand and can facilitate evaluation of colour specification schemes.

Colour is an increasingly affordable and common feature of graphics displays. As a result, both the computer graphics programmer and user find themselves plagued with the problem of colour selection. Two main obstacles complicate this task. The first is that the tools available for specifying colour are incompatible with traditional human conceptions, for instance, it is not natural to think of 'yellow' as 100% red plus 100% green plus 0% blue, as is necessary with a colour raster system. Even the colour specification systems [Smith 1978], developed to overcome this problem, may be restrictive in their oversimplification of human colour perception and classification. Secondly, there are often disparities between the objective reality of a display and its perceived appearance. Visual effects occur which can make a scene distasteful, irritating or cause it to be misinterpreted. In order to understand, predict and control these effects it is desirable to familiarize ourselves with the principles of colour vision and to develop a 'feel' for colour use. Since in many respects the aesthetics of colour remains subjective, users must learn to rely on their own intuition with regards to the effectiveness of the displays they create. The purpose of the software we have developed is thus twofold: to provide the user with a means by which he can visually appreciate phenomena inherent in colour perception; and to provide a tool with which various colour specification systems can be evaluated.

In a graphics system, the excited phosphor deposits which create an image on the CRT (Cathode Ray Tube) emit light which impinges on the viewer's eyes. Receptors in the eyes convert this light into neural impulses which carry signals to the cortex of the brain where a percept or thought is formulated. However, the act of seeing is not a passive, photographic task. Rather, vision encompasses an active procedure whereby the sensory data received is encoded, compressed, and examined. Thus, the viewer's final perception depends on these various transformations that the data undergoes as it travels from its inception at the eye, through the visual pathway, to a level of consciousness. Our study of colour vision will therefore comprise an examination of the physiology of this data path: at each stage of the process the intricate mechanisms used will be discussed along with any perceptual effects they may cause. Furthermore, since the physiological schema is not completely understood, we are forced to examine some subjective colour phenomena from a psychophysical or stimulus-effect point of view (this is where visual experience is crucial to understanding the effects). However, both the physiological and psychophysical results will prove to support the basic principle of operation of the visual system: that selectivity and comparison are used to derive meaning from the incident visual data [Bloomer 1976]. This postulate will help us understand the machinations involved in colour vision and hence to better cope with colour selection.

This work was supported by the Natural Science and Engineering Research Council of Canada under NSERC grant No. A3022. The paper was typeset by Waterloo Computer Typography on an APS-μ5 using software developed by Richard J. Beach.

PHYSIOLOGY OF THE EYE

The first stage in data compression and transmission occurs in the eye. The eyeball is optically constructed to filter both the quantity and content of the incoming electromagnetic radiation.

The main parts of the eye are shown in figure 1. Light enters the eye via the transparent membrane called the cornea. The cornea is actually the main focusing unit since its boundary with air presents the only surface with a substantial difference in refractive index [Gregory 1977]. In order for light to reach

the inner part of the eye, however, it must pass through the small opening or pupil, which is regulated by the pigmented iris. The iris controls the amount of light energy entering the eye, constricting or dilating depending on light intensity. Behind the pupil is the lens which acts as a fine focusing unit. The shape of the lens is variable to accommodate for object distance by changing its radius of curvature. The cornea and lens operate together to focus the light onto the sensitive retina, which is essentially composed of layers of nerve cells and covers approximately 200 degrees of the rear inner circumference of the eyeball [Haber 1980]. It is in the receptors of the retina that the light energy will be converted to neural impulses.

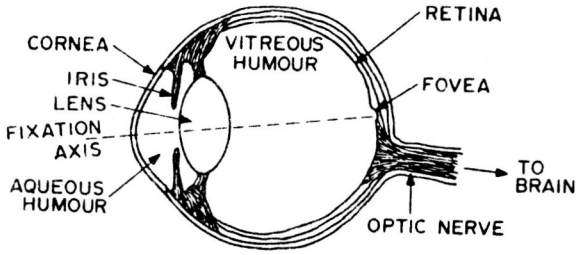

Figure 1 – Cross-section of the human eye [Padgham and Saunders 1975]

Recall that light is a form of electromagnetic radiation. Newton's prism experiment led to the discovery of the wave property of light when he showed that white light is refracted differentially by a prism to reveal its component wavelengths - the colours of the rainbow, running from red through yellow, green, and blue to violet. Colour thus corresponds to light containing narrower ranges of wavelengths than white light. The absorbent characteristics of the cornea, lens, and eye fluids restrict the sensitivity of the eye to radiation of wavelengths between 400 and 760 nanometers, which is therefore known as the visible spectrum. The cornea and the lens selectively absorb blue light, especially any radiation with wavelength less than about 400 nanometers. The aqueous and vitreous humours, fluid masses which help maintain the shape of the eye, also absorb considerable amounts of light energy particularly at longer wavelengths [Padgham and Saunders 1975]. Because refraction varies with wavelength, a certain amount of chromatic aberration occurs in the eye. The cornea and lens will refract shorter wavelength light more than longer wavelength light. This suggests that the eye cannot focus greatly differing wavelengths simultaneously which can cause difficulties when trying to focus on two distant forms, one red and the other violet. The violet one will appear considerably blurred while the red one is in focus [Boynton 1979]. Since the cornea and lens are not perfect optical instruments, and because considerable absorption and light scattering occurs in the ocular media (cornea, lens, aqueous humour, and vitreous humour) only a small fraction of the radiation incident on the cornea actually gets as far as the retina and that which does is not well focused [Haber 1980].

THE RETINA

The retina, with its complex and highly organized system of nearly 200 million neurons (nerve cells) performs the initial neural processing of light. Here incoming light patterns are detected by receptor cells and important elements of the visual field are accentuated as the resulting neural signals are transmitted through the nerve cell layers so that only relevant information is encoded for distribution to higher centers in the brain.

Retinal Structure

There are three main layers of cells in the retina: the receptor layer, the bipolar layer, and the ganglion layer (see figure 2). Many intricate interconnections exist between and within these layers. Light absorption occurs by means of photochemical processes in the receptors. The cells can then pass on the signals by activating the next cells in the hierarchy via electrochemical processes. A signal will exit the eye from a ganglion cell whose long tail or axon forms part of the optic nerve connection to the brain. Amacrine and horizontal cells provide connections within the ganglion and bipolar layers respectively. These inter-layer cells have a very important role to play in pattern detection and the accentuation of certain elements of the visual field. A cross-section of the retina reveals that the photoreceptor layer is on the backside, behind the other layers. Thus, light must traverse the maze of ganglion, bipolar and other cells, as well as a web of blood vessels in order to reach the receptors. This implies a further reduction in the amount of light which reaches the receptors.

The retina contains two types of light sensitive receptors: rods and cones. The rods are responsible for what is called 'scotopic vision': our ability to see under conditions of low illumination. The cones are responsible for 'photopic' or daylight vision and for colour vision [Gregory 1977].

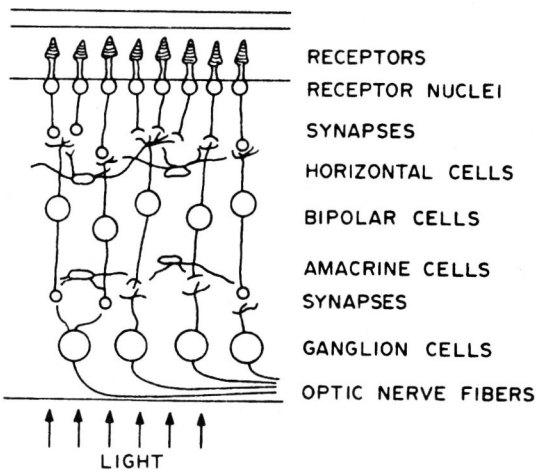

Figure 2 – Schematic diagram of a cross-section of the retina showing the cell layers [Padgham and Saunders 1975]

The most sensitive area of the retina is a depression at the back of the eye, only a couple of degrees wide, called the foveal pit. About 150 thousand of the 7 million cones in the retina are tightly packed into the 1 mm-square area of the central fovea which is devoid of rods [Boynton 1979]. There exists a one-to-one correspondence between these receptors and bipolar cells and between the bipolars and ganglion cells. This suggests high acuity in the fovea since elsewhere in the retina the connections are many-to-one. Furthermore, there are no blood vessels in the fovea to obstruct incident light and, because of the depression formed by the pit, less scattered light reaches the fovea so that it receives an image of higher resolution than elsewhere in the retina. The layout of the retina is thus designed to cater to our fixation point, occuring at the fovea, which is the area of greatest importance in the visual field.

The concentration of cones rapidly decreases outside the fovea while that of the rods increases to a maximum at approximately 16 degrees from the fovea and then also tapers off. The density of cones is constant and very low beyond 10 degrees. This suggests that colour detection is minimal in the retinal periphery. The rod density, however, decreases more slowly after it peaks so that in the peripheral area of the retina the rod concentration is much higher than that of the cones; there are approximately 120 million rods in the eye [Haber 1980]. Note that just below the fovea is located a blind spot - an area devoid of both rods and cones. This is where the ganglion cell axons are collected to form the optic nerve. Along with the decrease in receptor concentrations, the receptor to bipolar and bipolar to ganglion ratios increase beyond the fovea. Thus a bipolar in this area cannot retain the knowledge of which of its many receptors detected light . Spatial resolution will be very much reduced since a bipolar is responsible for a much larger retinal area [Haber 1980].

Receptors

As well as wave characteristics, light also possesses particle properties - it can be regarded as being composed of discrete energy packets or quanta called photons. Furthermore, the energy content of a photon is inversely related to its wavelength. That is, short wavelength light has photons with more energy than longer wavelength light. In order to be absorbed by a receptor in the eye, a photon must penetrate the receptor base, travel down its shaft to its outer segment, and there interact with the light sensitive chemical, the photopigment. When a photon 'isomerizes' a photopigment molecule, the molecule changes character - becomes bleached - and enables the occurrence of other chemical reactions which transmit the signal to the cells of the next layer. The photopigment molecule will eventually be regenerated and return to its original state, ready for another photon [Padgham and Saunders 1975].

The probability that a photon will be absorbed depends on its wavelength since the sensitivity of the receptors is a function of photon energy. Thus, receptors will absorb all wavelengths of light, but some with more likelihood than others. This probability distribution is described by the spectral sensitivity curves of the rods and cones, shown in figure 3. Rods are maximally sensitive to 500 nm (green) light. As illumination is increased, and we shift from rod (scotopic) to cone (photopic) vision, the spectral sensitivity curve of the eye shifts to the right. This is known as the Purkinje shift. Thus, under daylight conditions, the eye is most sensitive to 555 nm (yellow) light. The sensitivity decreases towards the edges of the visible spectrum. Since perceived brightness of light is related to the quantity of photons absorbed, the spectral sensitivity curves can predict the relative brightness of equal intensity lights of different wavelengths. Therefore, perceived brightness is not necessarily equal to physical brightness, or luminance. This is supported by the observation that two differently coloured patches, say red and yellow, will not appear to be equally bright even though their measured luminances are equal - the yellow will appear brighter under daylight conditions [Gregory 1978].

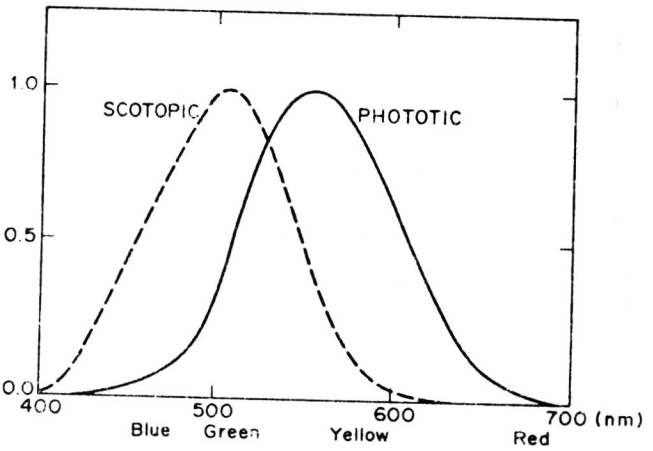

Figure 3 – Photopic (cone) and scotopic (rod) spectral sensitivity curves. [Wyszecki and Stiles 1967]

The rods and cones are distinguished by their shapes and photopigments. The rods are cylindrical and contain the photopigment called rhodopsin. One rod cell contains an average of 10 thousand rhodopsin molecules. Only one photon is needed to isomerize one rhodopsin molecule. The cones, so called because of their conical form, behave as the rods do in absorbing a photon. However, three different cone types have been discovered in the retina, each with its own distinct photopigment and spectral sensitivity curve. Each cone type is sensitive to all wavelengths of light but favours a different section of the visible spectrum: short wavelength (S) cones are most sensitive to short wavelength light while the medium (M) and long (L) wavelength cones are maximally sensitive to medium and long wavelength light respectively (see figure 4). Note that the sensitivity to the blue or short wavelength region is much less than that to the longer wavelength green and red regions. This is related to the fact that there are fewer blue or S cones than M or L cones in the retina and in the fovea it can safely be assumed that there are none [Boynton 1979]. This is

one of the factors contributing to our poor form perception of blue objects - a blue shape displayed on a CRT screen will appear to have a blurred contour.

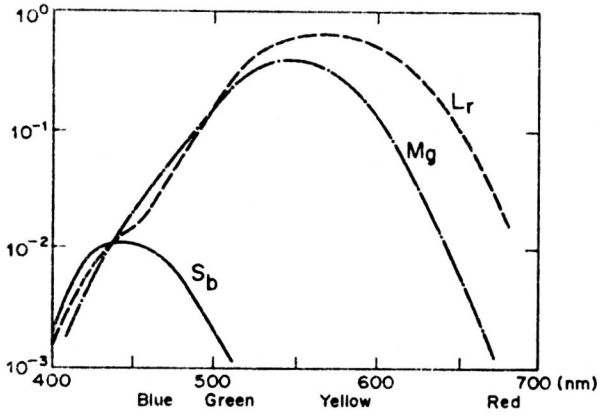

Figure 4 – Spectral Sensitivity curves of the 3 cone types. [Walraven 1974]

Although rod and cone sensitivities vary with wavelength, the effect of absorbing one photon does not. This principle of univariance states that once an isomeric reaction is initiated, all of the wavelength and energy data associated with the offending photon is lost. This implies that the rods and cones by themselves are colour-blind. The colour data will only be reflected in the relative activities of the three cone types which must be compared at higher levels in order to encode the colour information. Because of this we can effectively enable the perception of any colour by simply varying the intensities of concurrent red, green, and blue lights. This is the principle used to create a colour picture on the face of a CRT. Phosphor dots on its surface will emit light if they are excited from behind by a beam of electrons. The light intensity is related to the level of phosphor excitation and thus to electron beam strength. Dots of three phosphor types, one glowing red, one green, and one blue, in close enough proximity to appear concurrent, will then be able to produce (to within the limitations of the phosphor) any possible colour. Thus, to produce a colour on the CRT, we only need to specify the red, green, and blue phosphor excitation levels required.

One drawback of the CRT colour specification method is that we do not usually think of colours in terms of their component red, green, and blue intensities, but rather in terms of their hue (the name of the colour) and its variations, e.g. light-blue, dark-green, yellow, etc. Colours are thus easier to work with conceptually if an 'appearance' specification scheme is used. With this in mind alternate schemes have been developed. For example, Smith (1978) presents a method for mapping between the red, green, and blue intensity levels, and three psychological colour descriptors, those of hue (H), saturation (S) (the degree of a colour's departure from gray), and value (V) (the amount of a colour's departure from black, i.e. an approximation to brightness). Although it is easier to select colours by defining H, S, and V values, this system overlooks both luminance and perceived brightness for the sake of simplicity and incorrectly suggests that white and fully saturated blue, green, yellow, and red are equally bright. The photopic sensitivity curve, as we have seen, says that we do not perceive them as such. To partially correct for this problem Smith presents an alternate model in which value is replaced by a dimension he calls brightness (L). In this scheme the luminance is taken into account by using for L an approximation to the intensity of light emitted by the CRT phosphor triads. Again, however perceived brightness is ignored. One scheme which does consider both actual and perceived brightness is the NTSC (National Television System Committee) YIQ-standard which weights the red, green, and blue values unequally to partially compensate for perceived brightness. Such schemes do facilitate colour specification, but the user should also be aware of their limitations.

Information Processing in the Retina

In order for photon absorption to be signalled in the brain, the news of receptor excitation must be propagated through the neural network in the retina. However, we have seen that it is of little use to simply relay the receptor signals since they provide very little information on their own. In order to detect colour, form, location and other important visual cues, the output of several receptors must be compared and analyzed. The remaining retinal layers undertake to perform the initial phases of this task. Thus, the retina can be likened to a parallel processing system which uses much inter-processor communication.

Signal transmission is accomplished via chemicals called transmitters which are released by an excited neuron at its synapse (junction) with the next neuron. More than one sending neuron may converge on the receiving neuron which will only respond if enough of the excitatory chemical is released at the synapse to raise the electrical potential. The receptors provide a graded potential at their transmitting end which is a function of light intensity - the amount of chemical they release is proportional to photon absorption. All the cells beyond the receptors transmit signals by using action potentials - they always fire the same amount of transmitter so that their level of excitation is indicated by their firing rate.

Besides excitatory connections, inhibitory connections also exist. These inject a counter-excitatory transmitter at the synapse thereby reducing the electrical potential. Thus, the net potential difference at the junction will determine the response of the next cell. All neurons have a resting or spontaneous rate of transmitter discharge so that the synaptic potential can actually be driven below its non-excited level [Padgham and Saunders 1975].

Recall that signal flow is from receptors to bipolar cells to ganglion cells and out the optic nerve. If groups of receptors

have excitatory connections with bipolars, the rate of bipolar response will be proportional to the level of photon absorption in its associated receptors. Such excitation pooling can be continued at the ganglion level so that receptor excitation is spatially summed. This activity will cause signal amplification as more of the retinal area, which feeds the ganglion, is absorbing quanta. The effect of inhibitory connections is the opposite. Suppose receptors A and B are connected via excitatory synapses to bipolars A and B (see figure 5), which are in turn connected via excitatory and inhibitory connections respectively, to a ganglion cell. Stimulation of both receptors will excite both bipolar cells. Then, bipolar A will attempt to excite the ganglion while bipolar B will reduce A's effectiveness. This process is called lateral inhibition and causes signal attenuation. Both excitation and inhibition are functions of the sending cell activity, although a greater threshold potential must be overcome for inhibitory transmitter release.

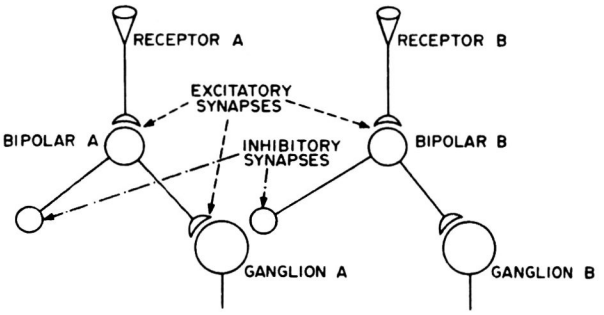

Figure 5 – Excitatory and Inhibitory connections between bipolar cells and ganglion cells. [Haber 1980]

Excitation and inhibition are both present at most synapses. It has been found that horizontal and amacrine cells provide inhibitory and excitatory connections among the cells of a layer. This implies that intra-layer interactions also exist with the possibility of feedback connections. The result of this complex networking is that "output from a particular receptor produces differential effects in the neurons of the retina depending on the number of quanta absorbed and which other receptors are absorbing quanta" [Haber 1980 p.67].

THE BRAIN

The brain performs the final stages of processing on the encoded data sent to it by the eyes. Although many of the mechanisms used in the brain are not as well understood as those of the eye, others have been shown to be extensions of those used in the retina.

Signals emerging from each eye travel along the optic nerve fibers which merge at the midline of the brain. Then the fibers are divided so that signals from the right side of each eye continue on to the left hemisphere of the brain and those from the left sides continue on to the right hemisphere. In this way data from the eyes can be compared. Before reaching the cortex, where a perception will be finally formulated, the majority of the signals travel through the lateral geniculate nucleus (LGN) of the thalamus. The LGN appears to be a simple relay station: although it may undertake some data processing using feedback from the cortex, studies have shown that cells in the LGN react to stimulus in the same way ganglion cells do. Hence, we will discuss LGN and ganglion cell activity together [Boynton 1979].

LGN

We have seen how spatial summation and lateral inhibition are used to accumulate receptor activity over a retinal area and then funnel the data to higher neural levels. The net result is that the LGN cells, like the ganglion cells, are each responsive to a certain retinal area. This area is called the receptive field (RF) of the cell and activity in the receptors of the RF will trigger some sort of action, either excitatory (increase in cell firing) or inhibitory (decrease in firing), in the cell. Due to the complex retinal connections, LGN and ganglion cell response is dependent on the pattern and/or wavelength of the stimulus on the RF. Two types of cells have been distinguished: spatially opponent (spectrally nonopponent) cells which do not recognize colour but respond to luminance patterns, and spectrally opponent cells which do recognize colour. [DeValois and DeValois 1975].

Two types of spatially opponent cells have been found. One, which has been dubbed the +Wh-Bl cell, has excitatory connections to a small central region of its RF and inhibitory connections to an annular surround (see figure 6). Such a cell is then maximally stimulated by a stimulus covering the center of the RF with the surround remaining dark. If the light stimulus is moved to any spot in the surround or if the size of the central stimulus increases into the surround the response of the cell will decrease. If instead, the center of the RF is dark and the surround is illuminated, the cell is found to be in a state of maximal inhibition (its firing rate falling much below the spontaneous level). Spatially opponent cells with the opposite behaviour have been observed. These are +Bl-Wh cells which have an inhibitory RF center and an excitatory surround. A central dark spot surrounded by illumination will maximally excite these cells while a central stimulus in a dark surround will inhibit them (see figure 6). The spatially opponent cells behave similarly to light of any wavelength and thus are also referred to as spectrally nonopponent cells. This is believed to be because the same cone types feed both the center and the surround of the RF giving it a fairly uniform spectral sensitivity [DeValois and DeValois 1975; Boynton 1979].

One of the main purposes of the spectrally nonopponent cell organization is to enhance visual edges. This is demonstrated by the Mach band effect, a phenomenon which is noticed often in computer graphics. If the visual field contains an edge separating dark and light regions, a darker band will be perceived on the dark side of the edge, while a lighter band will appear on the light side of the edge. By examining the

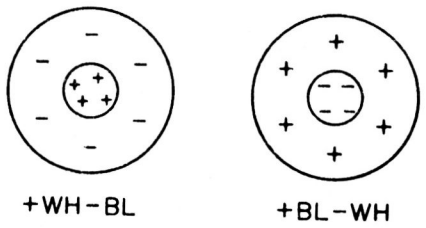

Figure 6 – Maps showing the effect of a stimulus on RF areas. [DeValois and DeValois 1975]

spectrally nonopponent cell reaction to edges we find a physiological reason for this to occur. To simplify the discussion, imagine the edge sweeping across the RF of a +Wh-Bl cell (see figure 7). When in the dark region the cell experiences no stimulation and is in its resting state. As the edge enters the RF, light encroaches on the inhibitory surround, thus driving the cell into inhibition. Once the edge moves into the RF center some excitation will be introduced which will counteract the inhibition. Since central excitation overpowers inhibition from the surround, the cell will reach a peak of excitation when its center is completely illuminated. Afterwards, as more of the surround is lit, more inhibition will be introduced again decreasing cell activity. When the RF is completely illuminated both excitation and inhibition exist and even though the excitation dominates slightly, the cell activity will not be very much above that of its resting state. The case of a fixed edge can be examined by considering all such RFs in the vicinity of the edge, many of which may overlap. The response curve obtained by summing LGN/ganglion cell activity for these RFs will be similar to the one obtained above. Hence, we see a darker band next to the edge in the dark region and a lighter band next to the edge in the light region. The response of +Bl-Wh cells will have the same visual effect except that their response curve will be reversed. Thus, the eye operates to accentuate luminance edges and borders which are important visual data [Haber 1980; Fiorentini 1972; Ratliff 1972].

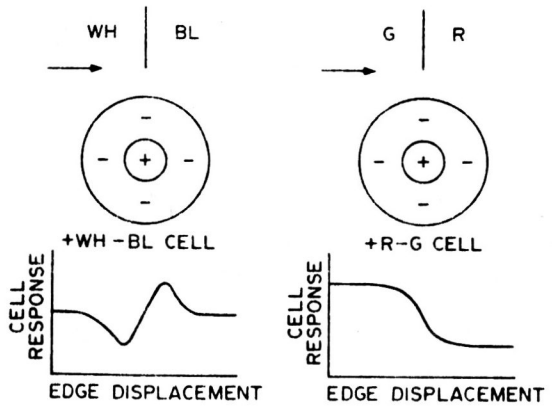

Figure 7 – Schematic diagram showing how opponent cell response varies as a visual edge sweeps across its RF. [Haber 1975]

Spectrally nonopponent cells constitute only one third of the LGN and ganglion cells. The rest do respond differentially to colour. It is the discovery and study of these spectrally opponent cells which has provided the evidence necessary to show that colour encoding is performed at levels beyond the receptors.

Spectrally Opponent Cells and Colour Vision

Early colour theorists recognized that the number of colour receptor types in the eye must be limited. It did not seem plausible that there were specialized receptors for each of the many colours we can distinguish. Such a proliferation of receptors would mean serious degradation of visual acuity under coloured light, due to a low density of responding cells, while acuity in white light would be high. This is in fact not the case. Not only can we see as well in coloured light as in white, but visual acuity is reduced in white light due to chromatic aberration effects [DeValois and DeValois 1975]. Thus, the two early theories of colour vision postulated that there exist a small number of colour receptor types in the eye and that between them they determine the wavelength (colour) of the stimulus. Although they are brilliant deductions, these theories have one major pitfall: they attempt to confer the entire task of colour determination on the receptors while neglecting the possibility of encoding at higher neural levels.

Young and Helmholtz expounded the trichromatic theory of colour vision which states that there are three types of colour sensitive receptors in the eye: one for red, one for green, and one for blue; and furthermore, that all colours are seen by various levels of simultaneous excitation of these three receptor types. This theory, as we have seen, is supported by the fact that there are three types of cones in the eye with different spectral sensitivity curves. However, the theory fails to indicate how the excitation levels in each of the receptor types are monitored and compared. Neither does it explain why we never see reddish-green or yellowish-blue, nor why simultaneous excitation by red and green produces the sensation of yellow, nor why we recognize four fundamental, distinct colours: red, green, yellow, and blue. In this way, it does not explain how colour is encoded. A subsequent theory presented by Hering, attempts to answer some of the questions ignored by trichromacy. Hering's opponent-process theory also claims that there are three types of colour receptors in the eye: a red-green receptor, a yellow-blue receptor, and a black-white receptor; but that each cell responds in one of two opposing and mutually exclusive ways depending on the wavelength of the stimulus. Thus, a black-white cell would be excited by white but inhibited by black, a red-green cell excited by red but inhibited by green, and a yellow-blue cell excited by yellow but inhibited by blue. This does seem to explain why we recognize four fundamental colours, and why we never see a red-green or a yellow-blue [Jameson and Hurvich 1957]. It also partially explains afterimages, another phenomenon unaccounted for by trichromacy. If we stare at a colour (say red) for approximately one minute, and then shift our gaze to a neutral gray surface we will see the complement (in this case green) of the colour

we had been looking at previously (yellow and blue are another complementary pair). This afterimage effect can be explained if we consider the inhibitory-excitatory behavior presented by Hering. It has been proven that at most synapses in the brain a surge of inhibitory chemicals (neural rebound) will occur upon removal of a prolonged stimulus. Thus, when we stare at red for a while and then remove the stimulus, the red-green cell will be thrown temporarily into inhibition causing a green afterimage to be perceived. From our discussion of the univariance principle, it should be immediately obvious that Hering's theory cannot be valid at the receptor level: each receptor alone cannot determine wavelength. However, by its opponent excitatory-inhibitory nature, the theory may be a valid description of cell response at higher levels. The LGN/ganglion spatially opponent system already discussed greatly resembles Hering's black-white cell. Furthermore, the spectrally opponent cells also discovered correspond in part to his yellow-blue and red-green cells. From these theories and discoveries the modern model of colour vision has been derived: the three cone types at the receptor level (the S, M, and L cones) feed, via additive and subtractive interconnections, the LGN and ganglion opponent cells. It is in these higher-level cells that the comparitive processing of the retina is expressed in the variation of cell activity with wavelength, i.e. colour encoding.

Work by DeValois and associates [DeValois and DeValois 1975] has disclosed the existence of four spectrally opponent LGN/ganglion cell types. The response of these cells depends on the colour of the stimulus incident on the receptors in their RFs; they are excited by some wavelengths and inhibited by others. For instance, the red-excitatory green-inhibitory cells (+R-G) are maximally excited by a red stimulus in their RF and maximally inhibited by a green one. Cells with the opposite reaction are just as common, these are the green-excitatory red-inhibitory cells (+G-R). Together these two cell types form the RG system. A YB system also exists consisting of yellow-excitatory blue-inhibitory cells (+Y-B) and of blue-excitatory yellow-inhibitory cells (+B-Y). These react to yellow and to blue stimuli [DeValois and DeValois 1975].

It has been discovered that most spectrally opponent cells also have a spatially organized RF. This means that they can act as spatially opponent cells to luminance patterns and as spectrally-opponent cells to colour. We can think of the RF map of such cells as the sum of the spectral and spatial RF maps (see figure 8). Specific cone types feed the RF center with excitatory connections while the surround is fed by different cone types via inhibitory connections (e.g. a +R-G cell has excitatory red cone connections to its RF center and inhibitory green connections to its surround). An interesting result of such an RF organization is that it does not provide an edge enhancement mechanism for colour boundaries across which there is no luminance change. Suppose a red-green boundary is moved across the RF of a +R-G cell and that the red and green regions are of equal luminance. (Recall that the M and L cone sensitivity curves overlap,thus we are speaking of the net effect of replacing a red stimulus by a green one.) Initially,

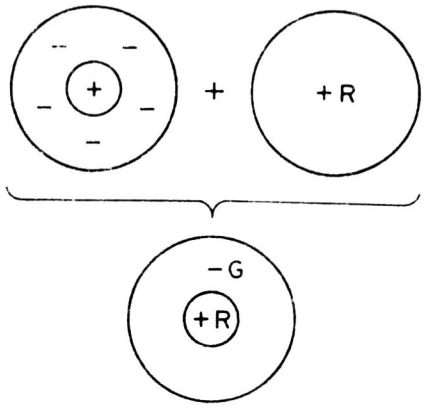

Figure 8 – RF map of a spectrally and spatially opponent cell as a sum of spatial and spectral RF maps. [DeValois and DeValois 1975]

the +R-G cell will be maximally excited since its RF will be completely inundated by red light. As the edge enters the RF surround the green begins to inhibit cell activity. Once the edge enters the RF center, inhibition will simply increase as red-excitation to the center is removed and as green-inhibition to the surround is increased. This continues until the entire RF sees only green; the cell is then in its maximally inhibited state. There is thus a smooth transition between excitation and inhibition. Although there is no chromatic edge enhancement at this neural level, recent evidence [Ware and Cowan 1981] suggests that it is performed but at higher levels in the cortex. Here double opponent cells have been discovered which have RFs with red-excitatory green-inhibitory centers and green-excitatory red-inhibitory surrounds (as well as mirror image cells). Such cells could theoretically produce chromatic Mach bands (e.g. red appearing redder in a green surround than in a red one) in a manner similar to the +Wh-Bl cells.

Spectrally opponent cells operate by comparing receptor activity and therefore must receive signals from at least two cone types - cones with different spectral sensitivities. The colour data is then obtained by differencing the activity levels of the cones of each type. In the case of the RG system, the cell inputs are from the medium and long wavelength cones. This is supported by the fact that if the L and M cone spectral sensitivity functions are differenced, two new curves are obtained with peaks at 640 nm, which is what we see as red, and at 500 nm, which we see as green. (In the case of +R-G spatially opponent cell, L cones are considered to feed the RF center while M cones feed the RF surround; the connections would be reversed for a +G-R cell) [DeValois and DeValois 1975]. The receptor connections to the YB system of cells are still in dispute: some researchers claim that S and L cones feed the YB cells, others claim that S and M cones do. The most popular hypothesis is that both M and L cones as well as S cones provide information to the YB cells: the yellow input being derived from the summed activity of the L and M cones

and the blue input from the S cones [Boynton 1981]. This certainly provides an explanation for the fact that red and green lights, when summed, form yellow. Figure 9 schematically shows these cell connections.

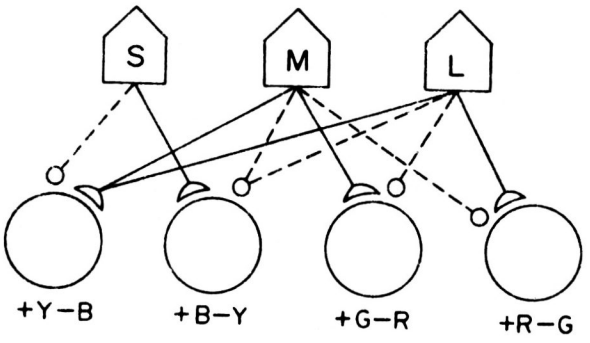

Figure 9 – Receptor connections to spectrally opponent cells. [Boynton 1979]

We have seen that LGN and ganglion cells encode luminance data as well as colour information. Since the brightness of a stimulus is echoed by the level of receptor activity, luminance measures are obtained by summing receptor outputs [DeValois and DeValois 1975]. However, it has been discovered that S cones (blue cones) do not contribute. Because of this and the fact that they are so sparsely distributed, especially in the fovea, S cones do not contribute much to contour perception nor is our visual acuity of blue detail very high [Boynton 1979].

Thus, the function of the LGN and ganglion opponent cells is to encode and compress receptor data, funnelling it into channels destined for the cortex of the brain. Two types of channels are recognized: the luminance channel sums the L and M cone outputs delivering brightness data to the cortex; the opponent-colour channels deliver colour data - the R-G channel feeding from L and M cones and the Y-B channel from all three cone types (see figure 10) [DeValois and DeValois 1975]. Further processing will then be undertaken by higher cortical centers.

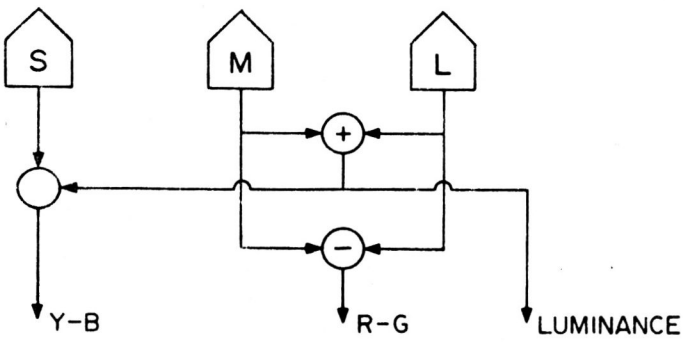

Figure 10 – Receptor inputs to visual channels. [Boynton 1979]

Visual Cortex

From the LGN nerve fibers travel to an area of the cortex in the brain's occipital lobe called the visual cortex [Boynton 1979]. Although neural activity here has been much less researched and is thus less understood, it appears that the comparative parallel processing of the retina is simply continued in several cortical layers.

If we recall the activity of an LGN opponent cell with both spectral and spatial characteristics, say a +R-G cell, it responds with maximal excitation to both a large red stimulus inundating the entire RF and to a small white stimulus incident only on the RF center. From the activity of this cell alone we are not able to distinguish between the two stimuli. However, if the output of this cell were compared with that of a +Wh-Bl cell with an overlapping RF, we would be able to make the distinction. A task of the visual cortex must therefore be to compare LGN output. The cortex is also topologically organized but the RFs corresponding to its cells will be larger than those of the LGN cells. The patterns of light and colour, incident in the RFs, which excite the cortical cells, are much more complex - some cells are excited by light lines on dark backgrounds, others by moving edges, others by specific hues, etc [Boynton 1979]. Thus the increase in function specificity and in RF size is simply continued in the neural layers of the visual cortex [Gregory 1978]. The final perception is considered to be derived from the output of many cortical cells, each of which provides a different characteristic of the visual scene [Padgham and Saunders 1975].

SUBJECTIVE COLOUR PHENOMENA AND FIRST HAND EXPERIENCE

We have seen that throughout the visual process comparison is used to encode data. It should not come as a surprise then that the appearance of a colour is influenced by the colour and luminance of its surround. [Itten 1961; Albers 1971; Truckenbrod 1981] However, much of what is understood about the physiology of colour perception only affects very small regions of the visual field and cannot account for large scale interactions occuring over extensive areas of a visual scene. For instance, the size of opponent cell receptor fields is too small to explain "simultaneous contrast" effects such as the reddish appearance of a neutral gray patch embedded in a green surround, or the impression of relative depth ordering which is produced by viewing yellow, red and blue colour patches against a black background. Higher, undiscovered cortical activity must be responsible for such phenomena. Psychological factors may also play a part in that some colours may convey "warmth" and "joy" while others "cold" and "desolation". Even though psychophysicists have defined some general principles based on individuals' subjective responses to physical stimuli (while ignoring the physiological contributions) such visual effects can only be fully appreciated by experiencing colour interaction and expression first hand.

A raster graphics system offers a powerful and flexible environment in which to experiment with all these phenomena. We have implemented a simple but effective program (in C on a PDP 11/45 running Unix) with which to investigate such effects. The user is allowed to specify an arbitrary number of rectangles (typically between 2 and 5) which are drawn on the screen of a 512 by 512 by 24-bit frame buffer in the order selected (see figure 11). The colour of each region may be individually controlled by manipulating sliders with a tablet and a puck (after selecting the HSV button). Any given rectangle may be grown or shrunk about its center. The colours associated with any two regions may be interchanged (using the FLIP button), or may be forced to be identical and altered simultaneously via the sliders (using the 2HSV button).

The program may be easily used to set up displays illustrating the effect of various surrounds on lightness and size, simultaneous contrast, border contrast, colour balance, various kinds of colour harmony, and so on. For example, the effect of surround on apparent size is seen in figure 11. A white square embedded in black seems to be larger than a black square of the same size embedded in white. Some phenomena are demonstrated much more forcefully when the characteristics of the surround are altered interactively than they are by several distinct static displays (as might be constructed from coloured paper). For example, simultaneous contrast can be experienced by setting up a medium gray square within another gray one. If the gray surround is gradually changed to red, the inner gray square obtains an increasingly greenish tint. Luminance contrast can be similarly experienced by placing a red square within a white one. As the white surround is interactively changed through the grays to black, the red square becomes increasingly brighter. Many of the colour exercises outlined by artists [Itten 1961; Albers 1971] can also be quickly performed. We have in fact used the program to construct a tutorial sequence of some 30 slides illustrating contrast effects discussed by Itten [Itten 1961]. The package can also be used to test other colour specification systems by replacing the HSV to RGB conversion routine. In this way we can become more aware of the limitations of the system being used and its suitability to the application(s) for which it will be used.

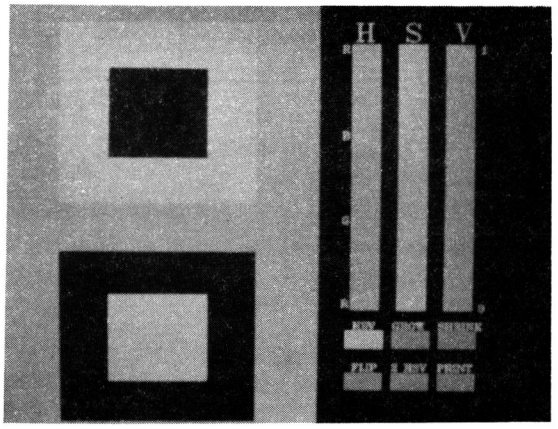

Figure 11 – Colour test-bed program at work showing the effect of surround on size: the white square in the black surround appears larger than the black square of the same size in the white surround.

CONCLUSION

Colour vision is a complex, not completely understood process, but what is known provides insight into understanding, predicting and controlling colour effects in computer graphics. We have seen that comparison and contrast appear to be central themes of perception mechanisms but only by visually experiencing colour effects can they be fully appreciated.

REFERENCES

[1] Albers, Josef, *Interaction of Colour*, Yale University Press, (London: 1971).

[2] Bloomer, Carolyn, *Principles of Visual Perception*, Van Nostrand Reinhold Company, (New York: 1976).

[3] Boynton, Robert M., *Human Colour Vision*, Holt, Rinehart and Winston, (New York: 1979).

[4] DeValois, Russell L. and Karen K. DeValois, "Neural Coding of Color", *Handbook of Perception*, **5**, (Seeing), Academic Press, Edward C. Carterette and Morton P. Friedman, ed., 1971.

[5] Fiorentini, A., "Mach Band Phenomena", *Handbook of Sensory Physiology*, **7**, part 4 (Visual Psychophysics) Antrum, Jung, Loewenstein, McKay, and Teuber, ed., Springer-Verlag, (New York: 1972).

[6] Gregory, Richard L., *Eye and Brain – the Psychology of Seeing*, McGraw-Hill Paperbacks, (Toronto: 1978).

[7] Haber, Ralph Norman and Maurice Hershenson, *The Psychology of Visual Perception*, Holt, Rinehart and Winston, (New York: 1980).

[8] Hurvich, Leo M. and Dorothea Jameson, "An Opponent-Process Theory of Color Vision", *Psychological Review*, **64**, *6*, 1957, pp. 384–404.

[9] Itten, Johannes, *The Art of Color*, Reinhold Publishing Corporation, (New York: 1961).

[10] Padgham, C.A. and J.E. Saunders, *The Perception of Light and Colour*, G. Bell and Sons Ltd., (London: 1975).

[11] Ratliff, Floyd, "Contour and Contrast", *Scientific American*, March, 1972, pp. 91–101.

[12] Smith, Alvy Ray, "Color Gamut Transform Pairs", *Computer Graphics (Proc. Siggraph '78)*, **12**, *3*, 1978, pp. 12–19

[13] Truckenbrod, Joan R., "Effective Use of Color in Computer Graphics," *Computer Graphics (Proc. Siggraph '81)*, **15**, *3*, Aug. 1981, pp. 83–90.

[14] Walraven, P.L., "A Closer Look at the Tritanopic Convergence Point", *Vision Research*, **14**, 1974, pp. 1339–1343.

[15] Ware, Colin and William B. Cowan, "Chromatic Mach Bands: Behavioral Evidence for Lateral Inhibition in Color Vision", Technical Report, Optics Section, Division of Physics, N.R.C. Canada, Ottawa, Canada, (Submitted for publication).

[16] Wyszecki, Gunter and Stiles, W.S., *Color Science - Concepts and Methods, Quantitative Data and Formulas*, Wiley and Sons, 1967.