Excerpts from

PROCESSES IN BIOLOGICAL VISION:

including,

ELECTROCHEMISTRY OF THE NEURON

This material is excerpted from the full β-version of the text. The final printed version will be more concise due to further editing and economical constraints. A Table of Contents and an index are at the end of this paper.

James T. Fulton
Vision Concepts

jtfulton@neuronresearch.net

April 30, 2017

Copyright 2001 James T. Fulton

16 The Equations of Vision and their confirmation [1]

"The real progress is made when I let the data speak for itself"
A. Kimball Romney, Member NAS, 2005

This Chapter remains in editorial review at this time. However, the materials in Sections 16.1 through 16.4 are important and need to be available for comment.

16.1 Introduction

This chapter is designed to bridge the detailed discussions in the chapters of Part D and the wide variety of performance descriptors of biological vision developed in **Chapter 17**. It is primarily mathematical. This chapter will develop the equations needed to understand both the fundamental and the more subtle features of the visual system. It will also be useful in understanding many of the failure modes discussed in **Chapter 18**.

Lacking anything more than a conceptual model of vision, the community has been hobbled for the last half a century. The previous conceptual models have relied heavily on the concept of linearity. This has constrained the development of realistic equations and descriptors of vision. The equations of this chapter are designed to overcome the lack of a mathematical framework applicable to the visual process.

The very core feature of the visual system is its nonlinearity. This fact combined with its variability in amplification, on an individual spectral channel basis, are the keys to understanding the nominal operation of the system. To understand the more unusual features of the system, understanding the many asymmetries in the operation of the system is necessary. These asymmetries are primarily temporal in character.

Lacking a fundamental mathematical equation of vision, the community has fought over whether to use the linear laws of additive color mixing or subtractive color mixing in object space to explain the physiological sensing of color in perceptual space. However, the perception of the human visual system has been known to differ from that of photometric instrumentation employing linear signal manipulation for a long time.

The model developed here suggests that there is very little if any analog signal processing, as opposed to signal decoding and cognition, performed in the brain. In this context, signal processing deals with waveforms as a function of time where as cognition deals with more abstract symbols largely independent of time.

It is important that the subject of photometry and colorimetry be subdivided into the parameters of those fields related to object space and the parameters of those fields related to perception space. In this regime, perception is meant to be separate from cognition. Cognition may involve an interpretation of the perceived information. A very simple form of cognition has always caused difficulty in simple psychophysical experiments in colorimetry because the subject has difficulty in separating his response to brightness from his response to chromaticity.

A great deal of data has been collected relating to various major portions of the visual system. This data frequently exhibits very subtle features that are difficult for the individual investigator to even recognize, much less evaluate, in a proper context. The process is greatly aided by reference to a substantial model of the visual process. This chapter will examine a number of areas involving data of a comprehensive nature. It will also provide an interpretation of this data based on the overall model developed in this work.

---

[1]Released: April 30, 2017

A major conclusion resulting from **Section 16. 3.3** is that while the overall human eye is trichromatic, the retina of the human eye remains tetrachromatic, even in old age.

This chapter will also attempt to highlight key features of the data that should be watched for in future experiments and looked at very carefully. Thus, this chapter will emphasize the importance of good experimental design, careful data collection and subtle features of the results. Particular areas of review will include:

Invasive Experiments

+ invasive electrophysiological experiments involving the measurement of the Ganglion Cell outputs in response to luminous stimulation, i. e., cumulative responses through Stage 3.

Non-invasive Experiments

+ non-invasive electrophysiological experiments involving the ERG and similar techniques, basically through Stage 2.

+ non-invasive psychophysiological experiments designed to evaluate only a few parameters or situations, generally related to Stage 4 and beyond.

+ non-invasive psychophysiological experiments designed to aggravate information from more specific experiments in order to provide a comprehensive interpretation of the visual process. These experiments attempt to validate, and provide a method of validating a wide range of individual experiments.

16.1.1 Background

16.1.1.1 Terminology

16.1.1.1.1 An equation versus a function

In this and the next chapter, a series of descriptors will be used to describe the performance of the visual system in animals, and more particularly in man. Unfortunately, many of the descriptors found in the past literature have not been adequately defined. This section will review these archaic definitions and provide new recommended terminology.

To avoid undue confusion, this chapter will adopt a style where *named functions* will be functions of either time, wavelength or temperature. *Named equations* will involve unitary values resulting from the integration of all terms involving time, wavelength and temperature. The term function alone will also be used in the discussion in its generic form.

16.1.1.1.2 Illuminance versus irradiance

Because of the complicated nature of the radiation space to which the eye responds, it has always been difficult to provide a meaningful but short description of this environment. This has usually been done by assuming the eye exhibits a fixed or nearly fixed spectral response over a wide range of irradiant intensity almost regardless of the color temperature of the source. The result is a terminology based on the generic Latin word for light, the Lumen. Unfortunately, the eye does not exhibit a constant spectral response over the two regions normally considered. These are the imprecisely defined photopic and scotopic regions. In addition, the color temperature of the source of radiation is a crucial factor in determining the performance of the eye. This Chapter will demonstrate that the term illuminance is primarily an application oriented engineering term. It attempts to provide a scalar value for the integral of the irradiance of a given source, at a normally unspecified color temperature, over a nominal (essentially arbitrarily specified) spectral interval. The more fundamental scientific term is the irradiance. The irradiance is a very complex vector.

In scientific investigations, the problem becomes even more complex when the difference between the irradiance at the aperture of the eye is compared to the irradiance falling on the retina. It becomes still more complex when one attempts to use lumen based terminology when discussing the signal processing functions of the eye.

To understand the operation of the visual system at the research level, it is necessary to consider the

irradiance in quantal units (incident photons per unit of area per unit of time). The system does not respond to energy and the use of energy units are misleading.

16.1.1.1.3 Large signal versus small signal conditions

One of the major findings related to this work is the effect of the internal feedback within the adaptation amplifiers on the output signal level of the photoreceptor cells. They not only control the signal level within the neural layers of the retina, they maximize these signals within the dynamic range of the analog, electrotonic, neurons. Because of this action, all signals within the signal processing circuits of the retina must be considered large signals. The common tool of circuit analysis, assuming the small signal case (and therefore employing a linear approximation), does not apply to the circuits within the visual system.

16.1.1.1.4 Additive versus subtractive color mixing

pigment mixing versus sensation mixing

16.1.1.1.5 Photometry versus radiometry

In general, a photometer is not adequate for modern scientific investigation for three reasons. The absorption spectrum of the photometer is seldom appropriate to the psychophysical experiments involved, especially if those experiments are adequately planned. The actual absorption spectrum of the photometer is usually unspecified. The actual photodetector is frequently an energy sensitive device that does not give a response analogous to the quantum sensitive eye. A fourth reason is the false sense of security such an instrument gives to a novice investigator.

16.1.1.1.6 Luminous efficiency and “Quantum efficiency”

There is always interest in defining the "quantum efficiency" of the eye. This is readily done based on this model. The investigator need only consider a few specific characteristics.

Sometimes, it is not clear what is meant by the quantum efficiency. In the case of the eye, the quantum efficiency will be taken as a function of location on the retina. For each location, it will be taken as the number of electrical charges created in the outer segments of the photoreceptors at a specific location divided by the number of photons entering the pupil of the eye and directed at the above specific location .

This definition of quantum efficiency (based on engineering) is quite different from the definition of a quantum yield occasionally found in the literature and the threshold sensitivity (based on psychophysics). A definition of threshold sensitivity will be developed later in this Chapter.

The most important parameters affecting the quantum efficiency are;

+ the loss of imaging photons within the eye before they reach the focal plane of the optical system. These losses may be due to absorption, reflection and/or scattering.

+ the effective cross-section of the outer segments of the photoreceptors in the focal plane of the optical system. This cross-section is highly spectrally selective because of the matrix nature of the retina and the variable content of the matrix with respect to location, in terms of the type of photoreceptors present.

+ the effective absorption of the photons by the outer segments of the photoreceptors. This effective absorption is a strong function of the state of excitation of the outer segment.

The overall quantum efficiency cannot exceed 1.0 minus the loss in the optical system. This upper limit can be taken as the overall percentage transmission at a given wavelength. This upper limit can often be approximated by the species specific transmission of the main lens group, see Figure 2.3.2-1 in Chapter 2.

The effective absorption cross-section must be calculated for each spectral type of photodetector. The forbidden bands of the energy band diagram for each spectral type of photodetector prevent a short, medium or long wavelength detector from absorbing photons outside its specific spectral band. There is

also space between the various photodetectors. Therefore, as a first approximation, the peak spectral quantum efficiency of the eye is limited in each absorption region by the total cross-sectional area of each type photodetector per unit area (as a function of position in the retina).

[Does the following make sense or is there a different/better modeling method]

If a long, medium or short wavelength photon enters the outer segment of an appropriate photoreceptor, and the outer segment is not in a highly excited state (i.e., it is dark adapted), there is almost a 100% probability that it will be absorbed. This is true irrespective of its exact wavelength because the photon will encounter over 1000 absorbing surfaces in the space frame of chromophores in the outer segment.

[May need revision]
Therefore, the quantum efficiency of the eye, under dark adapted conditions, is given by the product of the optical transmission of the optical system times the effective cross-section of the outer segments for each spectral type of photoreceptor.
sed by input photon excitation and

16.1.1.2 Glossary

In the following glossary, many terms will be described as *Archaic* for scientific purposes and an alternative term suggested. These *Archaic* terms have frequently found broad application in engineering and need not be considered obsolete in that regime. However, their use in state of the art scientific investigations should be avoided unless the requisite auxiliary descriptors are provided. The intent in this work is to include the traditional lumen based terminology frequently in this Chapter but to concentrate on the more precise terminology only in Chapter 17.

16.1.1.2.1 Terms independent of spectral composition

Stiles-Crawford Effect– The result of introducing irradiance to the retina of an eye along a non axial ray. It represents the variation in intensity actually recorded using the elliptical optics of the eye compared to that expected by assuming the eye employs spherical optics. This effect is now recognized to include two distinct aspects, the first kind and the second kind (**Section 17.3.7**).

Lumen based terms

Absolute luminance threshold--Only applicable to the scotopic luminous intensity region and scotopic absorption spectrum associated with that region. Refers ultimately to the threshold level of the ganglion cells associated with the luminance channel of the signal processing circuitry of the human eye. Frequently obscured in psychophysical experiments by the quantum noise associated with the incident irradiation in each photoreceptor channel.

Lumen--

Luminance discrimination function--Technically a description of the change in intensity required to reach a given threshold, either perception or recognition related, as a function of wavelength. Usually determined using an irradiation source and narrow band filters. This source is usually unrelated to the lumen based irradiation regime and is frequently of a poorly specified color temperature. Usually, only the illumination calibration equipment is lumen based. **[Check this paragraph]** See XXX

Chrominance discrimination function— See Chromatic discrimination function

Chromatic discrimination function

Luminosity--

Luminosity function, Photopic, C.I.E.--*Archaic* for s. p.

Luminosity function, Scotopic, C.I.E.--*Archaic* for s. p.

Luminous efficiency function

C.I.E terms based directly or indirectly on Lumen based units

Chromaticity Diagram, C.I.E.---*Archaic* for s. p. Most recently modified in 19XX but based on inadequate mathematical approximations with regard to the operation of the actual eye. See Chromaticity Diagram for Research, New.

Photopic luminosity function, C.I.E.--*Archaic* for s. p.

Scotopic luminosity function, C.I.E.--*Archaic* for s. p.

Uniform Chromaticity Scale Diagram (introduced by Judd)--*Archaic* for s. p. **see pg 301 in O Society book.**

Radiance based terms

Chromaticity Diagram for Research, New--

Frequency response of the eye, spectral--A state variable critically dependent on the adaptation state of the individual photoreceptor channels of the eye and the spectral content of the irradiation employed in experimentation.

Frequency response of the eye, temporal--A state variable critically dependent on the adaptation state of the individual photoreceptor channels of the eye. Evaluation requires specific attention to the intensity and spectral content of the irradiation employed.

Transient response of the eye, compete--

16.1.2 Graphical presentations 101

There has recently been a dialogue in the vision literature that has highlighted the question of conformality between data sets[2]. This is a somewhat esoteric subject usually reserved for cartographers and mathematics majors. However, its achievement in graphical presentations is equal in importance to that of the proper use of statistics. The lack of conformality, as in the C.I.E (1931) Chromaticity Diagram, leads to precision in the absence of accuracy. One may read coordinate points off of this Diagram but one may not read angles or project straight lines as tangents to various curves. Tangent lines are not straight lines in nonconformal graphs such as the above Diagram.

The geometrical problems with the C.I.E. (1931) Diagram have been known for a long time. Ellipticity in a presentation associated with threshold data (in any field) usually suggests a lack of conformality in the presentation. The ellipses MacAdam (1942) overlaid on the CIE 1931 chromaticity diagram provide an example[3]. They suggest the significant lack of conformality in the underlying diagram. The origin of the problem is illustrated in Wyszecki & Stiles (1982). Farnsworth (1944) recognized this lack and led an empirical attempt to make the CIE 1931 chromaticity diagram conformal. It culminated in a nearly orthogonal (but not conformal) CIE 1960 UCS diagram and a nearly conformal CIE 1976 UCS diagram.

The problem is one long known to cartographers. How does one present a true map when projecting a spherical surface onto a flat surface. Truth can be specified in terms of the correct area of the features projected or in terms of the correct distances (and a variety of other criteria). In general, both area and distance cannot be preserved in a map representing a portion of a globe. The goal of such projections is to maintain conformality. Conformality requires two conditions. The preservation of perpendicularity between any set of lines on the new projection representing perpendicular lines on the original is the first. This condition is called orthogonality. The second condition is that the length and direction of line segments will be scaled properly in the new projection. This condition is known as equi-angularity. Equi-

---

[2]Sankeralli, M. & Mullen, K. (2001) Letter to Editor, *Vision Research*, vol. 41, no. 1, pp 53-55

[3]MacAdam, D. (1942) Visual sensitivities to color differences in daylight. J. Opt. Soc. Am. Vol. 32, pg. 247

angularity generally requires that the unit vectors associated with two lines at right angles must be in the same proportion in the new projection as in the original. The word “original” is used here for a reason. We want to be able to speak of the “original” as either a prior graph or as a prior dataset based on a uniform number line. The uniform number line of mathematics is the series of cardinal numbers in order and extending from minus infinity to plus infinity. Any data set that one normally collects is based on a uniform number line and is therefore conformal when plotted with any other data set so collected. The challenge is to present these data sets on a conformal graph. The best known conformal graph is generally called a Cartesian coordinate system. It uses two orthogonal axes with uniformly spaced number lines associated with the two axes. Data plotted on such a graph is considered conformal in the first instance. Projection of the plotted dataset and axes onto any other graph may or may not be conformal.

16.1.2.1 The problem of a color pigment equilateral triangle

Investigators since Young have attempted to plot colors on an equilateral triangle. While artistically pleasing, the format is quite limiting. When shown, scales are usually drawn with the individual levels drawn parallel to all three edges of the triangle. This format does not represent the data orthogonally and it is difficult to manipulate the data graphically. Only the grouping of data points can be shown on such a coordinate system. When the corners are specified to represent the pigments or chromophores of vision, the color space does not allow for the perception of color at wavelengths beyond the peak wavelengths of the chromophores. Only occasionally can relationships be shown meaningfully. Goldsmith (1990, pg 311) noted: the representation of such pigment triangles do not represent quantitatively the metrics of color space as determined psychophyically. The graph space cannot be transferred conformally to an orthogonal graph space.

16.1.2.2 Extension of the color triangle to an equilateral tetrahedron

With the recognition that most animals exhibit tetrachromatic vision, several investigators have attempted to move from a three cornered triangle representing the chromophores of vision to a tetrahedron. When used as a graph space, the structure consists of four nonconformal surfaces and the volume enclosed by the surfaces remains nonconformal. As above, the grouping of data points can be shown but no unique relationships can be discerned from such a presentation.

16.1.2.3 The problem of the C.I.E (1931) Chromaticity Diagram (as an example)

The C.I.E (1931) Chromaticity Diagram was the result of an intensive empirically based series of experiments by several different laboratories in the 1910's and 1920's. The instrumentation of the day was crude by the standards of today. The empirical base was flawed. It assumed the visual system employed linear additive color processing based on three absorption chromophores (or chromatic channels). This caused an immediate problem. Every effort to deconvolve the datasets into three independent components as functions of wavelength invariably required one of the components to exhibit a negative absorption at some wavelength. This theoretically impossible condition was glossed over by defining an equivalent (non-realizable) “Standard Observer.” This observer employed the absolute value of the resolved components as a function of wavelength in order to eliminate the theoretical error. Further manipulations were introduced to convert the data set three parameter dataset from a quasi-equilateral one to a right-equilateral one that could be plotted onto a Cartesian coordinate system. These manipulations are traced in Wyszecki & Stiles[4].

The essence of the above manipulations was the highly nonconformal representation of a 2-dimensional color space we know as the C.I.E. (1931) Chromaticity diagram. After the presentation of this Diagram, it was overlaid with many functions attempting to explain its meaning and showing other features. The classic overlay was the attempt to present a uniform number line, represented by the spectral locus, on the Diagram. The resulting curve was no longer a straight line and its spectral intervals were no longer equally spaced. These two facts illustrate the degree of nonconformality present in the diagram. The degree of nonconformality is also suggested by the ellipses of MacAdam.

Farnsworth (1944) recognized the above lack of conformality and led an empirical attempt to make the CIE

---

[4]Wyszecki G. & Stiles, W. (1982) Color Science, 2nd Ed. NY: John Wiley & Sons, pp 121-138

1931 chromaticity diagram conforma[5]l. It culminated in a nearly orthogonal (but not conformal) CIE 1960 UCS diagram and a nearly conformal CIE 1976 UCS diagram. The degree to which he succeeded is illustrated in **Section 17.3.3.4.1**.

16.1.2.4 Introduction of a conformal color space

Boynton began exploring color vision from a psychophysical perspective in the 1950's and soon developed a new color space. This color space was based more on the Hering concept of color differences in an algebraic space instead of the Young-Helmholtz concept that suggested color summation. This color space was orthogonal but it remained largely conceptual and various scales were associated with the axes by different investigators. A particular problem was the lack of attention to the color temperature of the light sources used in laboratory experiments. This made it difficult to correlate results unless the investigators agreed to standardize their conditions (whatever those conditions were). The conditions adopted frequently introduced parameters into the experiments that were not expressly understood.

This work introduces an expandable conformal space compatible with and providing absolute scales for the Boynton and more recent psychophysical color spaces. The basic space is formed by using the spectral locus from 300 nm to 655 nm (extendable to 1100 nm when appropriate) as a uniform number line. The uniform number line is folded, using multiple 90 degree angles, to form several conformal spaces as shown in **Figure 16.1.2-1** [xxx following text requires more complete figure. caption is a mismatch] Each of these has a different application in vison research. In (**B**), a “cube” is formed that can be used to describe the color performance of a tetrachromat. The space would form a true cube if it was truncated at the peak wavelength of each of the chromophores, i. e., 342, 437, 532 and 625 nm. However, this smaller color space would not accommodate the larger perceivable color space of vision. The full space is technically a right parallelepiped. This form demonstrates unambiguously how the perception of "white" is achieved (See **Section 17.3**). (**C**) shows a degenerate form of the color cube that provides a 2-dimensional color space compatible with the conventional interpretation of a trichromat. This space extends beyond the spectral peak at 437 nm in order to accommodate the perceptual color space extending to approximately 400 nm. This extension from 437 to 400 nm is not continuous with the lower part of the figure. While both sections are conformal, there is no conformality across the 437 nm line. Note the difference in coloration at wavelengths less than 437 nm. A constant color is shown horizontally for each wavelength. (**D**) shows pseudo-color cube that is easier to work with. It folds the spectral locus at 437 nm but keeps it *in-plane*. It applies to the complete tetrachromat and shows its complete perceptual color space on a flat surface. As in (**C**), the left and right sections of the figure are conformal. However, there is no conformality across the vertical spectral line. “White” cannot be shown explicitly as a single point in this representation. The fold points of 437, 532 and 625 nm will be justified in Chapter 17.

[5]Farnsworth, D. (1944) The Farnsworth rectilinear uniform chromaticity scale diagram No 38. Memorandum Report # 44-1 New London CT: Med. Res. Lab. U.S. Submarine Base *or* Wyszecki & Stiles, 2nd Ed. pg. 311

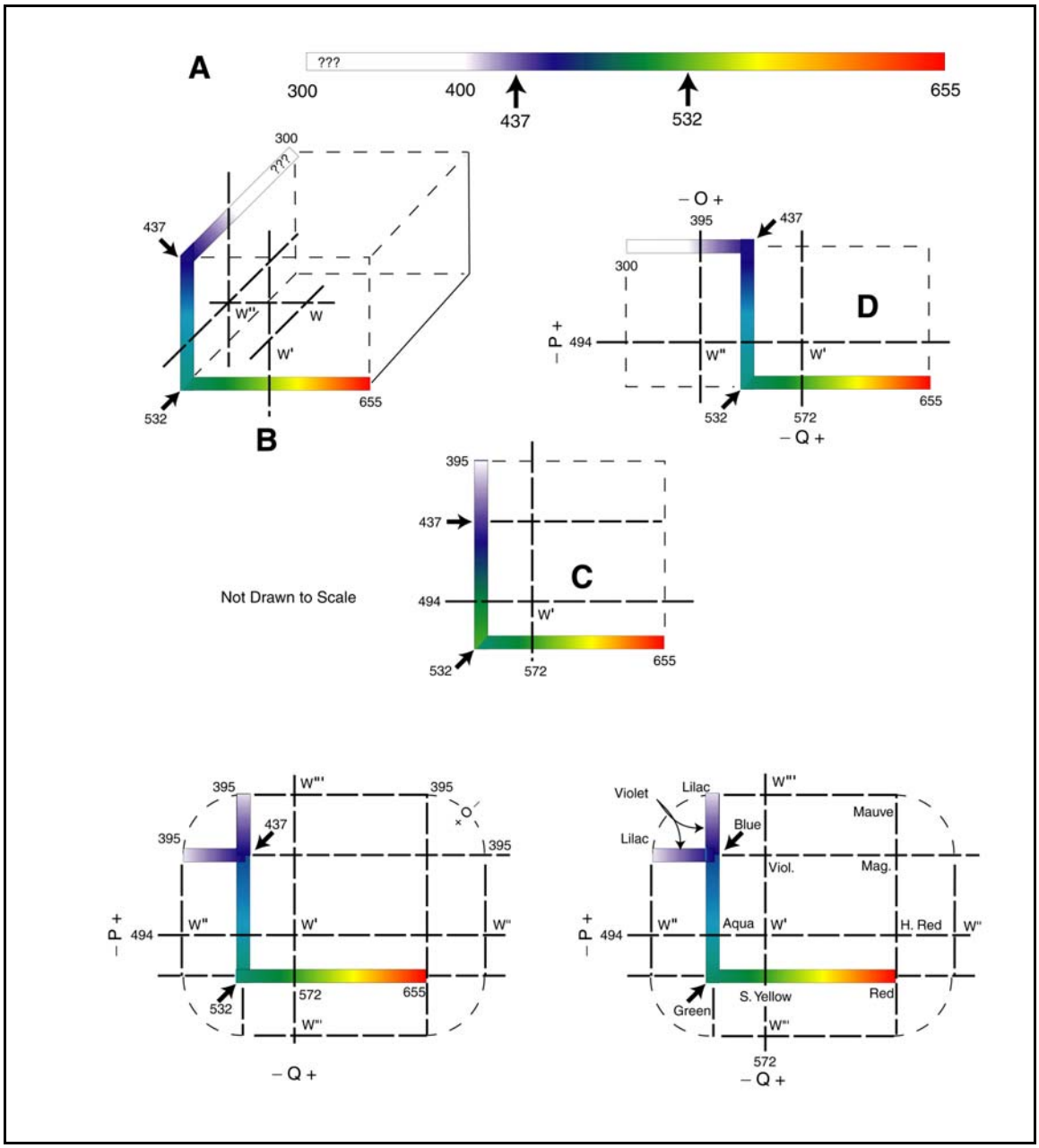


Figure 16.1.2-1 Formation of candidate conformal color spaces. A; the spectral locus from 300 to 655 nm used as a number line. B; the spectral locus bent at 532 nm to form a plane and bent again out of that plane at 437 nm to form a "color cube." C; the spectral locus bent at 532 nm as above but not bent at 437 nm but truncated at 400 nm. D; the spectral locus bent at 532 nm to form a plane and bent again at 437 nm but staying in-plane.

16.1.2.4.1 Basis for conformal transformations used in chromaticity diagrams

While it is possible to prepare a conformal graphical representation of the color space of vision by plotting the independent functions associated with the O–, P– and Q–channels, it is not obvious that a continuous spectral locus can be associated with these scales.

From a mathematical perspective, it is possible to treat the spectral locus of the electromagnetic spectrum as a continuous number line, whether plotted with respect to frequency or wavelength. Assume the spectral line is continuous in the region of 300 to 1000 nm. It is possible to fold such a line by employing a mathematical treatment known as a conformational transformation.

The simple conformal transformation between a single linear axis (represented by the conventional spectral intensity plot) and an orthogonal x-y plot is given by the equation $w = z^{\frac{1}{2}}$ (z to the ½ power). Using this transformation and polar coordinates, it is seen that a linear scale in the linear z plane is folded at a specified point. This results in two linear scales in the w plane radiating orthogonally from the apex formed at the fold point[6]. There remains the question of where the fold point should be. Although the discrimination characteristics of each chromatic channel extend beyond the nominal center wavelength of the M-channel chromophore, it appears 532 nm. is the appropriate fold point. If the absorption spectra of the M-channel was narrowband monochromatic, 532 nm. would clearly be the only rational fold point. A similar transform can be used again to create a second 90 degree bend. Let this bend occur at 437 nm. If the second bend is out of the plane formed by the first bend, the resultant graphical space is three-dimensional.

It is the above mathematical treatment that is used to justify the displays of the spectral locus as found in the chromaticity diagrams of Section 17.3.3.

16.1.2.5 The consequence of changing from a wavelength to a frequency baseline

The recent introductory text by Rodieck presents the subject of vision based on spectral frequency rather than wavelength[7]. While this change may be well-intended, the consequences are significant. His discussion of the choice between energy and photon flux and the conversion from wavelength to frequency on page 513 to 515 is interesting but not substantive. He does not address conformality at all. The transfer of information acquired along a linear wavelength number line does not transfer to a linear frequency number line conformally. These two quantities are reciprocal and the relation between them is hyperbolic. As a result straight lines in one graphic space will not be straight in the other. Angles between two lines in one space will not be the same in the other. Any graphs based on the historical vision database acquired with respect to wavelength will require reinterpretation.

16.1.3 Introduction of feedback theory

Investigators in the field of vision have largely ignored the discipline of feedback theory until the arrival on the scene of Tranchina[8]. His team was probably the first to introduce the mathematics of feedback theory, even if limited to elementary concepts. By associating several of the tools of electrical engineering with the problem of understanding the conceptual aspects of feedback theory, Tranchina has moved the field forward. In doing this, he correctly defined the number and type of circuit elements required as a minimum between the photoreceptor cells and the horizontal cells. However, he noted that further progress would require a more detailed understanding of the physiology and morphology involved. It will also require the introduction of more advanced techniques.

Tranchina employed a number of reductive techniques of circuit theory to determine the approximate form of the feedback used in the early stages of vision circuitry. However, he did so on a global basis that did not separate the signal path into individual circuits. He was only able to define a class of appropriate feedback models. Thus, his approach showed conceptual feasibility at a first (gross) level. He did review the features of several of these models. However, he did not determine the applicable boundary conditions required to further isolate the correct model. Each solution within his class contained two linear filters and one static nonlinear element. Some solutions included an element to multiply two signals. Neither of the models presented included forward gain within the feedback loop. He did not justify his definition of the filters as linear. However, this is a normal initial assumption. He also did not concern himself with the

---

[6]Kraus, J. (1953) Electromagnetics. NY: McGraw-Hill Pp. 541-547

[7]Rodieck, R. (1998) The First Steps in Seeing. Sunderland, MA: Sinauer Assoc. Inc. Pg 515

[8]Tranchina, D. & Peskin, C. (1988) Light adaptation in the turtle retina: embedding a parametric family of linear models in a single nonlinear model. *Vis. Neurosci*. vol. 1, pp 339-348

greatly decreased transport velocity of electrons within an electrolytic medium at biological temperatures. This velocity is typically 4400 meters/sec instead of the nominal 300,000,000 meters/sec of light and the typical 150 to 200 million meters/sec velocity of unconstrained electrons in a metallic conductor. The resultant factor of about 50,000 leads to a significant transport delay over distances measured in microns and millimeters. This transport delay must be considered when reporting the total delay of signals at the horizontal cells.

Just as in the case of the P/D Equation, it is necessary to subdivide the photosensing model between excitation and the output of the horizontal cells more completely and to determine the boundary conditions more precisely in order to proceed. This requires two things. Additional wisdom (additional training plus additional experience in the field) to be brought to bear on the problem and the introduction of a process called circuit realization. The latter is a mathematical procedure to determine if a circuit can actually be built, and if so, the simplest circuit form it may take.

The introduction of more advanced techniques would allow the selection of more precise configurations actually involved in vision. This work has provided the necessary additional details and provided a more precise circuit model of the feedback found in the early stages of the visual process. It does employ techniques beyond those introduced by Tranchina. These include,

+ the realization that the reference signal Tranchini sought may be internal and applied at the output of the feedback circuit rather than be related to the input signal.

+ the use of complex algebra to account for the finite time delays related to the transport velocity of electrons in quantum-mechanical and electrolytic circuits,

+ the introduction of the concept of internal feedback as a complement to the more readily obvious techniques of external feedback and

+ the introduction of multiplicative circuit elements not envisioned by Tranchina.

It also explicitly recognizes the critical fact that the visual system is a change detector due to a zero in the numerator of the transfer equation of the adaptation amplifier. This zero is a result of the feedback mechanism used in vision.

16.1.3.1 The requirement for Z--plane analysis when stage 3 circuits are involved

The analytical tools used by Tranchina are designed for use in an analog signal environment. When analyzing pulse signals (the action potentials) of stage 3, it is necessary to recognize the fact that the system becomes a sampled data system in the process of encoding by the ganglion cells. A special extension of S-plane analysis has been formulated to handle this class of signaling[9]. Z-plane techniques properly account for the signals in this environment. These techniques should be employed when attempting to describe the signals within or following stage 3 of the visual system.

16.2 Fundamental block and schematic diagrams of vision

The previous literature has not developed a comprehensive model of the visual system. This has prevented the community from advancing beyond the mathematical descriptions of Waddell and others. These have been based on the idea that the overall response of the visual system consists of a single signal that can be dexribed as the linear sum of signals from three spectral classes of photoreceptors. The assumption has been that the parameters of these photoreceptors are constant. The precise nature of these signals has not been specified as the electrical or chemical form or as to where they are measured.

With the significantly more detailed model of the visual system available from this work, it becomes possible to provide more comprehensive, and accurate, descriptors of the visual process. These new descriptors are both mathematical and graphic in style. The equations of this chapter are based entirely on the electrolytic signaling capabilities of the visual system. No chemical signaling mechanisms are provided or allowed. Furthermore, the photoreceptors are shown to exhibit a variety of parameters that are variables. Their output signal is a voltage at the pedicle of each cell. The model makes the basic assumption that the signals at the output of the photoreceptors are combined in a variety of fundamentally independent signaling channels.

---

[9]Truxal, J. (1955) Automatic Feedback Control System Synthesis. NY: McGraw-Hill, pp 500-611

16.2.1 The architecture of the visual system

For continuity and to place the descriptors in context, the overall block diagram and the schematic diagram of the fundamental signaling path will be repeated here. Most of the other pertinent diagrams will be described by their location elsewhere in this work. The subdivision of the overall visual system into the stages introduced in **Section 1.5.1.2.2** is particularly important. It allows the system to be analyzed in a piece-meal manner. It also allows different mathematical techniques to be introduced when describing different mechanisms.

To appreciate this chapter, the reader must have some familiarity with the calculus, complex number notation and either LaPlace or Fourier Transforms. Fourier Transform notation will be used in the general case. It is also important to note that Ohm's Law is not appropriate, and cannot be used effectively, in the study of vision. The inherent processes are not linear and require the use of the more general Kirchoff's Laws.

This chapter will not repeat the detailed calculations used to develop the Photoexcitation/De-excitation Equation and other detailed calculations. It will reference and assemble the results from these detailed calculations in forms that support the presentation of the overall descriptors of **Chapter 17**.

16.2.1.1 The Stage 1/Stage 2 interface

Understanding the interface between Stage 1 (Signal Detection) and Stage 2 (Signal processing) of the visual system is critical to understanding the operation of the visual system. The bulk of the visual system operates under constant signal level and large signal conditions where possible. It depends on the signal detection circuits of the photoreceptors to provide the necessary signal amplification and stabilization to satisfy these operating conditions.

As suspected for a long time, the key to understanding the operation of the photoreceptor cells is the fact that their individual outputs are logarithmically related to their input stimulus[10]. They also incorporate the adaptation amplifiers that are key to the apparent stability in the performance of the overall system. Such stability is not limited to color constancy but also includes the less obvious brightness constancy.

The interface between Stage 1 and Stage 2 is shown in **Figure 16.2.1-1** along with the lens of the physiological optics, Stage B. This figure will be discussed in detail in **Sections 17.2-17.4**. Its importance here is that it illustrates three crucial situations. First, the formation of multiple signaling channels at the output of the photoreceptor cells. Second the subdivision of the photoreceptor cells into their functional elements. Finally, a graphic representation of the signals carried by the individual channels at the location of the S-plane of the retina. This figure shows that the theory of this work is an extension of the earlier zone theory. The inputs to the photoreceptors consist of four spectral channels (instead of the previously assumed three) in accordance with Young-Helmholtz and the chromatic signaling channels follow the approach of Hering (except there are three chromatic opponent channels instead of two).

[10]Brindley, G. (1970) Physiology of the retina and visual pathway. 2nd ed. Baltimore, MD: Williams & Wilkins, pg 55

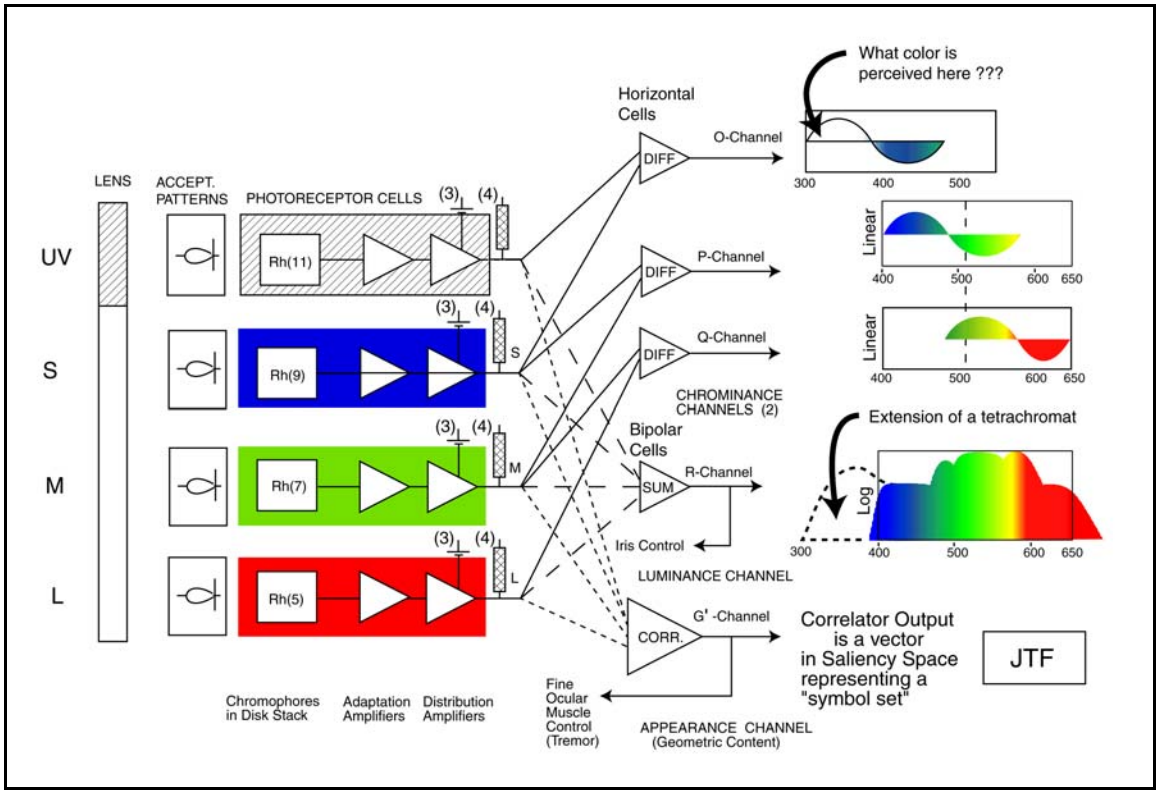


Figure 16.2.1-1 The luminance, chrominance and appearance channels of the eye of tetrachromats and aphakic humans. The spectral response in the O-, P- and Q- channels are shown as sinusoidal for illustration. The UV photoreceptor cells are known to be functional in humans of all ages. Research is ongoing to determine if the signal in the O-channel of the aphakic human is typical of tetrachromats. If it is, an aphakic human will be able to tell us what "color" other animals perceive in the ultraviolet.

Note carefully, there is no achromatic (or rod) sensor channel in this configuration. All of the necessary information is acquired from the spectrally selective sensor channels. This figure is similar to a figure in DeValois & DeValois[11] and shows a familial resemblance to many other figures[12].

16.2.1.2 Table of channel designations

The above figure includes a series of channel designations that are used throughout this work. However, it is not comprehensive. This figure has not addressed an additional series of appearance channels, the G-channels (or geometric channels) associated with the spatial characteristics of the retina. Only the G'-channel (alternately described as an S' channel in some figures because it involves a summation) of the appearance channels is shown. This particular subchannel is critical to the oculomotor functions of the eyes and thereby to the analysis of fine detail. It is related specifically to the foveola of the retina. Contrary to the view of DeValois & DeValois[13] who were studying the LGN) this channel does exhibit pure photoreceptor signals in the optic nerve between the retina and the brain. However, these paths go to the pretectum and not the LGN.

Chapters 16 & 17 do not address a majority of the G-channels of vision because they are less important in

---

[11]DeValois, R. & DeValois, K. (1975) Neural coding of color. *In* Carterette, E. & Friedman, M. *eds*. Handbook of Perception, Vol. 5, Seeing. NY: Academic Press, pp 117-166

[12]Uttal, W. (1981) A taxonomy of visual processes. Hillsdale, NJ: Lawrence Erlbaum Associates, pp 644-645

[13]DeValois, R. & DeValois, K. (1975) Op. Cit. pg 107 in Carterette & Friedman

human vision than in some other animals (particularly well studied in the cat). Their analysis is also more difficult because it is difficult to tease apart the individual mechanisms underlying the literature.

The middle column in the above figure contains three types of signal processing cells. The horizontal and bipolar cells are located in the retina. Their outputs are located in the S-plane. The cells associated with the G'channel are not located in the retina. They are located in the pretectum. They are an integral part of the POS. Whereas individual horizontal cells take differences between two spectral channels and individual bipolar cells sum the output from multiple spectral channels, the correlation process associated with the G'-channel path is more complex. This process involves a two dimensional array of cells performing a two-dimensional correlation over an area equivalent to that of the entire array of photoreceptors in the foveola.

It is desirable to provide clear labeling of the above channels to avoid confusion. The following labels will be used:

+ photodetection channels are labeled to agree with the peak spectral absorption region of their chromophore, the UV-, S-, M-, and L-channels. The subscript for the UV-channel will be abbreviated to U.

+ the luminance signal processing channel is labeled R–.

+ the chrominance signal processing channels are labeled O-, P-, and Q-, where
the O- channel involves the difference, U– minus S–,
the P- channel involves the difference, S– minus M–
the Q channel involves the difference, M– minus L–.

+ a polarization channel, generally based on a differential signal from two orthogonally oriented chrominance channels, will be described as an N–channel. Such a channel only appears in selected species.

+ a putative, and possibly non-existent channel within the brain, T-, characterized by a signal that will be described by T(E), the reconstructed chrominance vector used within the brain to completely describe the sensation of color.

+ the appearance signal processing channels take on a variety of forms in the animal kingdom. The general designation G-channel will be used. Alternately an A–channel designation may be used. When more specific areas are examined, the labels V-, W-, X- may be introduced. An additional special case applies to a number of families of *Chordata*, including the family *Hominidae*. These families have a well developed foveola that is part of a Precision Optical System involving a servomechanical loop designed to point the eyes very precisely. The appearance signals associated with this loop will be labeled G'. The appearance channels will not be discussed in detail in this Chapter. The POS is discussed in **Chapter 7** and other aspects of the G-channel are discussed in **Chapter 11.**

+ a "Y" channel has been encountered occasionally in the literature to describe a group of neurons with unidentified specific characteristics. In this work, it has been used to identify a group of spectrally different neural pathways leading from the stage 1 sensory neurons of the foveola directly to the PGN/pulvinar couple of stage 4 via individual stage 3 ganglion neurons.

+ An expression **Z**( ) will be used to represent a vector quantity in saliency space. This expression will include all of the (visual) information about a given scene known to the animal. This expression has no physical expression. It is a basic representation of the saliency map of the animal. It is limited to a partial description of the mental state of the animal. In the general context,

$$\mathbf{Z}( ) = a_0( ) + g a_1( ) + h a_2( ) + i a_3( ) + j a_4( ) + ... \qquad \text{Eq. 16.2.1-1}$$

Where the script letters represent imaginary operators describing a multi-dimensional space of currently unknown degree.

16.2.1.3 Location of major nonlinearities, asymmetries and gain variations

The above figure displays a variety of signal detection and signal processing channels that incorporate nonlinearities of various types. **Figure 16.2.1-2** illustrates these. The following paragraphs show the visual process in animals is **not** linear in any sense. The process includes many non-linearities. Any

hypothesis based on the assumption of linearity is bound to fail. An argument can be presented that the visual process is linear for small signals. However, the thresholding and signal manipulations found in the retina forcefully refute this claim, especially for the color aspects of vision.

The Greek letters shown along the bottom edge of the figure will be used to define the functional equations related to these signal planes. Plane A, or ρ_k, describes the physical plane between the disks of the outer segment and the dendrites of the neural portion of the photoreceptor cells. σ_k describes the base region of the first Activa within each photoreceptor cell after free electrons have been generated. Tk describes the output of the first Activa in its role as the adaptation amplifier. ν_k describes the current output of the second Activa in its role as a distribution amplifier. The voltage nodes, ω_k describe the voltages at the pedicles of the photoreceptor cells. The perception plane corresponds to the S-plane of Tomita and Svaetichin. It is proposed that the signals at this location properly describe the first order perceptions of vision in the biological system. These signals closely conform to the signals recovered by the stellate cells within the cortex. Significant differences only arise in the second order due to nonlinearities and asymmetries within the signal projection stage, stage 3.

The S-plane was named after Svaetichin who first recorded analog signals from within the retina in an organized way. The literature presents selected measurements taken at the S-plane of the retina. Many of these signals have been recorded using probes on the order of 1,000 times larger than the individual neural conductors. Therefore, the recorded signals must be viewed as a large set containing both individual significant waveforms as well as a wide variety of composite waveforms. Because of the variety of techniques used to measure and present these signals, an absolute determination of the signal polarity, in any individual or all animals, is not possible. Tomita describes the nature of the S-plane signals well[14].

The labels along the upper row of the figure describe the shorthand notation used to describe the signals along the circuitry. Their precise meaning will be defined in subsequent mathematics. In general, they are the amplitude of the voltage at the node in the presence of a standardized excitation. C_U describes the coefficient in front of the logarithm describing the output voltage of the ultraviolet channel photoreceptors. C_{OU}, or C'_U, describes the similar coefficient at the input to the Activa of the horizontal cells (in the case of the O–, P– & Q–channels) or the bipolar cells (in the case of the R–channel). C_{OH} describes the coefficient of the expression representing the output voltage of the horizontal cells of the appropriate channel. In the case of the R–channel, the coefficient would be C_{RB} to indicate the voltage at the output of a bipolar cell.

C_{OM} takes on a different meaning. There appears to be good correlation between the morphology of the ganglion cells and their physiology. It is the coefficient describing the pulse interval between action potentials in the O–channel at the output of the midget ganglion cells. In the case of the R–channel, the designation would be C_{RP} to indicate the coefficient describing the pulse interval between the action potentials in the R–channel at the output of the parasol ganglion cells.

---

[14]Tomita, T. (1965) Electrophysiological study of the mechanisms subserving color coding in the fish retina. Vol. XXX of the Cold Spring Harbor Symposia on Quantitative Biology. NY: Cold Spring Harbor Laboratory of Quantitative Biology pp. 559-566

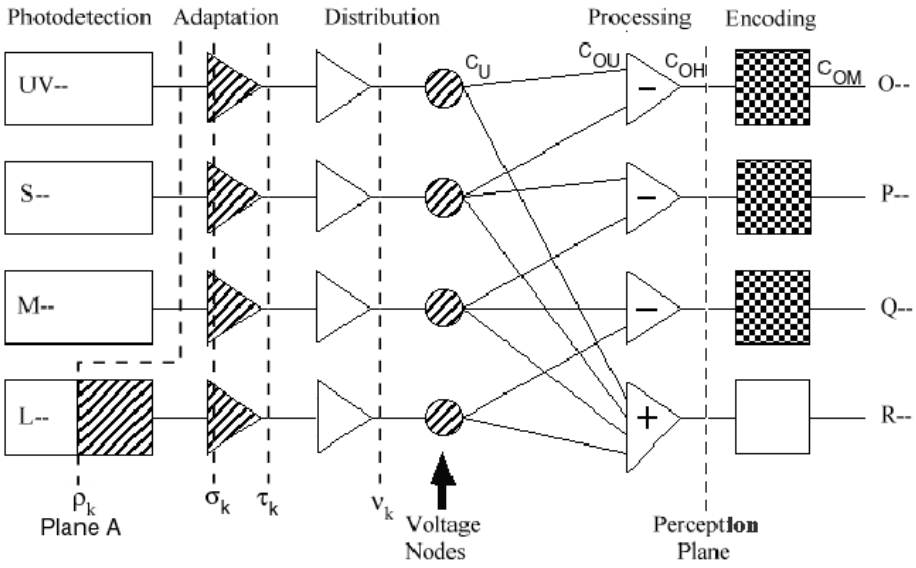


Figure 16.2.1-2 Important signaling channels, sources of non-linearity and asymmetry in the visual process of animals. Plane A corresponds to the equations of Wandell. The Perceptual Plane corresponds to the S-Plane of Svaetichin. The cross-hatched areas introduce non-linearities. The checkered areas introduce asymmetry. See text for details. This figure does not illustrate the appearance or G-channel. Human vision does not normally employ the UV-channel.

16.2.1.3.1 Nonlinearities in vision

The most striking non-linearity and photodetection level asymmetry involves the L-channel translation process. The transduction portion of the photodetection process (photons to excitons) is linear over many orders of magnitude in all of the chromophoric channels. This is not true in the translation portion (excitons to free electrons) of the photodetection process. The L-channel translation process involves a square law relationship in the transfer equation. This characteristic is the cause of the transition from the photopic to the scotopic spectral sensitivity function (luminance efficiency function) in humans and other animals exhibiting long wavelength spectral sensitivity.

The transfer function of the adaptation amplifier found in the photoreceptor cells of all animals is a stunning example of how far from linearity the visual system diverges. The mechanism underlying this transfer function involves an exponential mechanism with an argument that is between three and six. This mechanism is incorporated into an internal negative feedback loop. The resulting transfer function of the adaptation amplifier is a dynamic function of the current **input** to the amplifier from the photodetection process. This characteristic is the mechanism providing the very wide overall illumination sensitivity range of the eye.

The most fundamental non-linearity in the process of vision is at the output nodes of the distribution amplifiers, shown by circles. At this point, the current originally generated in proportion to the illumination creates a voltage by passing through a diode. The result is a voltage that is the logarithm of the current (and

intensity of stimulus). This voltage is sensed by the processing amplifiers that follow.

The signals from the processing amplifiers are encoded by the ganglion cells. The parasol ganglion cells create a series of action potentials in the R-channel whenever the incident signal exceeds a threshold. Any threshold constitutes a non-linearity in the mathematical sense. This threshold prevents the brain from receiving signals that are excessively noisy at low illumination levels. The midget ganglion cells create a series of action potentials under quiescent conditions and modulate that series of action potentials in response to the O--, P– & Q– signals. The modulation used is called phase modulation. The time between pulses (and not the frequency) is a direct function of the applied signal.

16.2.1.3.2 Asymmetries in vision

A process which is linear but defeats the additivity premise and therefore the superposition corollary is the spectrally selective subtraction occurring in the O-, P-, & Q- signal processing stages.

The encoding process in the O-, P-, and Q chrominance channels is required to handle an electrically bipolar signal. The midget ganglion cells perform the encoding without introducing a non-linearity with respect to signal amplitude. However, the technique used causes a significant asymmetry with respect to time among the signals decoded in the brain. Perception of the recovered information relating to the S-channel and the L-channel requires more time than does the perception of the recovered M-channel information. This asymmetry in perception time is the primary cause of the special effects related to flicker and other transient experiments.

16.2.1.3.3 Variations in amplifier gain in vision

A key feature in vision is the variation in adaptation amplifier gain associated with each photoreceptor. This variation is the key to color constancy and brightness constancy. Within the photopic region of the systems operation (a characteristic primarily of the system and only indirectly of the external light level), there is no perceived difference in scene brightness or color rendition. This region can extend over a luminous range of 3500:1 if the incident spectral content is not skewed too severely. The mechanisms providing this dynamic range are summarized in the next section.

An additional variation in gain is introduced by the P/D mechanisms at values of $\sigma \cdot F \cdot \tau < 1.00$. This gain term has been largely overlooked in the past. See **Section 16.4.1.1.1**.

16.2.1.4 Table of coefficients

Because of the complexity of the visual process from a mathematical perspective, it is useful to provide a brief summary of the coefficients used in the following equations. Each of the four channels of photodetection in a typical eye will exhibit a gain coefficient associated with the transfer function between the photons applied to a specific photoreceptor cell and the *current* appearing at the pedicle of that photoreceptor cell. This coefficient will be labeled K prior to its evaluation or assignment to a specific photodetection channel.

K-- When discussing the steady state performance of the visual system, K has the units of electrons per incident photon. It is only associated with points in the visual system prior to the logarithmic conversion of a current to a voltage. It can also be expressed in current divided by the rate of photons incident on the retina. In both the steady state and temporal performance cases, K may contain variables associated with the temperature of the animal, time and with the adaptation amplifier gain coefficient. K may also include a term related to the saturation effects in the conversion of photons into free electrons within the neural system. This expanded definition of K remains compatible with the conceptual K of the literature .

C– is the generic coefficient used to describe the ratio between the incident photon flux rate and a variable in the visual system following the pedicle of the photoreceptor cells. It is expressed as a voltage divided by the rate of incident photon flux. Alternately, it can be used to express the ratio between two voltages.

An additional form of C will be used when in the voltage domain. In accordance with the equation, ln C = C,' ln C' is a global constant related to the voltage term appearing at the voltage node, the pedicle of the photoreceptor cell.

Subscripts-- To avoid confusion, it is necessary to use subscripts to delineate the specific coefficients and

constants within the various visual signal processing channels. In the photodetection stages, only a single subscript is needed. The coefficients and constants will be; K_U, K_S, K_M, K_L, and C_U, C_S, C_M, C_L, and will be defined as above. The subscripts will relate to the UV-,. S-, M- and L-channels of photodetection.

When dealing with the luminance, chrominance, or appearance channels, it is necessary to employ double subscripts. The processing channel subscript will appear first. Thus, C_{RU}, C_{RS}, C_{RM}, C_{RL} will be used to describe the voltage found at the input to the Activa of the luminance channel summation circuit. An initial subscript of O, P, or Q will describe the voltage found at the input to the Activa of one of the chrominance differencing channels. Here again, the subscript O, designating the UV- minus the S-channel differencing circuit, does not normally apply to the atmospheric trichromats.

Because of the operation of the adaptation amplifiers in each photoreceptor cell, the constants related to the voltage at the pedicles of the photoreceptors, C_U, C_S, C_M, C_L, are all nominally equal. However, this is not true for the constants in the luminance signal processing channel, C_{RU}, C_{RS}, C_{RM}, C_{RL}. The corresponding constants in the chrominance channels appear to be nearly equal.

Additional notation may be introduced when discussing the encoding stage, related to the ganglion cells.

TABLE 16.2.1.4
TABLE OF COEFFICIENTS USED IN VISION

	AT	To				
	PEDICLE	O–HORIZ.	P–HORIZ	Q–HORIZ	R–BIPOLAR	G'–CORREL.
FROM						
Stage 1 pedicles	C_U	C_{OU}	C_{PU}	C_{QU}	C_{RU}	$C_{G'U}$
	C_S	C_{OS}	C_{PS}	C_{QS}	C_{RS}	$C_{G'S}$
	C_M	C_{OM}	C_{PM}	C_{QM}	C_{RM}	$C_{G'M}$
	C_L	C_{OL}	C_{PL}	C_{QL}	C_{RL}	$C_{G'L}$

16.2.1.5 Spectral parameters of the chromatic channels

In this Chapter, the spectral parameters of the individual chromophores derived in Section 5.5.11and tabulated in the Standardized Human Eye will be used. These parameters are those of the chromophores labeled the Rhodonines when in the liquid crystalline state at normal body temperature as found in the Outer Segments of the photoreceptor cells. In this state and applied to a linear substrate (opsin) as a coating, the Rhodonines exhibit an anisotropic absorption spectrum. This spectrum is not obtained by transverse measurements of the spectrum of individual photoreceptors nor is it obtained in measurements of dilute solutions. The family of four members exhibit essentially equal spacing across the visual spectrum from the UV to the Red. The parameters are those of a multilayer stack of this material as found in the visual system.

The above specifications are compatible with most of the spectra recorded in the electrophysiological environment..

The above specifications are compatible with recent micro-spectro-radiometric measurements of the in-vivo photoreceptors taken through the pupil. These measurements record the axial, anisotropic spectrum of the actual operational photoreceptors. Such measurements do not involve any assumptions about the signal paths of the visual system. They suffer primarily from the presence of extraneous material that can distort the spectral data, most frequently related to the L-channel chromophores because of hemoglobin in the blood and tissue.

16.2.1.5.1 Limitations on difference spectra

The above specifications are not compatible with the majority of the psychophysical literature related to the response of individual chromatic photoreceptor channels. This literature is based on psychophysical

experiments that attempt to decompose the luminosity functions into their components based on spectral differencing experiments employing differential adaptation and a linear model of the visual process. The most comprehensive recent study appears to be that of Stockman, et. al[15]. The second paragraph of that paper suggests why they continue to calculate the M- and L-channel spectral responses as very broad and overlapping to a very high degree while the S-channel, that is measured more directly is narrower and less overlapping.

One of the major problems with the spectral difference technique is the choice of adapting lights. Wald was able to isolate the S-channel chromophore almost completely using a yellow adapting light that suppressed both the M- and L-channel chromophores by more than a factor of 100:1. However, when he attempted to isolate the L-channel, he did not use the complementary color (see the New Chromaticity Diagram for Research). Instead of using azure, centered on 494 nm, he used blue. The blue light did not significantly suppress the M-channel. The result was a broad spectrum absent the S-channel but containing both the M- and L-channel.

Stockman, et. al. used a different approach. They employed a low level, one second duration, adapting light that appears insufficient to maximize chromatic adaptation based on this work. It was followed, cyclically, by a low level background light also of one second duration. Their figure 3 shows spectra that are very broad, contain significant response in the 400-475 nm region due to absorption by the S-channel, and overlap extensively. These spectra are easily decomposed further into their basic constituents using the model and mathematics of this work.

Difference spectra cannot be used to determine the spectra of individual chromophores in the absence of a sophisticated model of the complete visual system. The currently available estimates of the S-, M- and L-channel spectra based on psychophysics cannot be combined to produce the narrow peaks in the human luminosity functions related to the Brezold-Brucke Effect or the Purkinje Effect. On the other hand, they can be further decomposed into the underlying chromophoric spectrums.

16.2.1.6 The circuit architecture of the photoreceptor cell.

The electrolytic circuitry of the neural portion of the photoreceptor cell is the key to understanding the visual process. When coupled with the quantum mechanical mechanisms of the transduction process, these translation and amplification mechanisms define the dynamics of the visual process. A crucial element of this circuitry is the use of negative internal feedback within the adaptation amplifier of the cell. The mathematical description (although not the physical configuration) of this negative feedback circuitry is similar to that proposed by Tranchina (See **Section 16.1.3**). However, it does not include the “lead-lag” network he suggested. This type of network is associated with stage 2 signal processing but not that of stage 1.

As noted by Tranchina, it is necessary to analyze the actual physiology of the neural circuits, particularly of the photoreceptor cells, before determining the actual architecture of the stage. The actual circuitry of the photoreceptor cell differs from the class described by Tranchina in the following ways,

1. The “input filter” is comprised of the absorption due to the physiological optics combined with the P/D Equation describing the transduction mechanism. As opposed to his model, the P/D mechanism and the P/D Equation describing it are not linear.

2. There is forward gain within the feedback amplifier circuit.

3. The forward gain is controlled by a stabilizing circuit contained within the amplifier providing the forward gain described above.

4. The slowly changing component of the stimulus is treated differently from the fast changing component within the feedback amplifier circuit.

5. The overall forward gain of the complete circuit at zero frequency is zero. This is due to the high transfer efficiency of the feedback circuit at zero frequency. Chordates, including humans, are blind in the absence of change in the visual field.

---

[15]Stockman, A. MacLeod, D. & Johnson, N. (1993) Spectral sensitivities of the human cones. J. Opt. Soc. Am. A, vol. 10, no. 12, pp. 2491-2521

6. Feedback is obtained via an internal feedback mechanism. There is no physical path providing external feedback.

7. There is a significant delay at the output of the photoreceptor cell due to the finite transport velocity of electrons within the electrolytic medium of the cell (particularly in the extended axon segment).

Thus, his class of potential feedback models is revised to include,

1. An input filter that is nonlinear with signal amplitude.

2. A feedback loop that contains forward gain but no real forward filter element.

3. A low pass filter element in the feedback path.

When the signal from the feedback path is summed with the signal applied to the feedback stage, one result is the suppression of the DC component of the input signal. The signal at the output of the feedback circuit appears to have been passed through a high pass filter.

4. An amplifier providing the forward gain that is dynamically nonlinear based on a well known semiconductor breakdown mechanism. This mechanism controls the current gain of the Activa (the gain element). The result is a constant average current through the feedback amplifier regardless of the input stimulus (within the dynamic range of the feedback amplifier circuit).

5. A significant electrolytic transport delay between the distribution amplifier of the photoreceptor cell and the point of waveform measurement.

6. A static nonlinear element formed by the diode acting as a load impedance for the distribution amplifier of the photoreceptor cell.

This diode converts the current signal within the axon to the voltage signal sensed by the synapses and following neurons.

The resulting circuit architecture is shown in **Figure 16.2.1-3**. While this architecture is a member of the class defined by Tranchina & Peskin, it does not correlate well with the closest simile in their paper[16]. The transfer function of the P/D Equations associated with the disks and the Base of the Activa in each dendrite (microtubule) constitutes the filter labeled D(f) in Tranchina & Peskin. The diode forming the load of the distribution amplifier corresponds to their logarithmic element, h(z). Where they called for a filter in the forward path of the feedback circuit, this work supports a filter in the feedback path. The forward path contains only a wide bandwidth current amplifier. However, the effect of the low pass filter in the feedback path causes the overall feedback circuit to appear as a high pass filter with gain. The gain is shown as inversely proportional to the average input current. The feedback differencing circuit is shown within the operational amplifier symbol to emphasize that is not a physical element. It is an internal feature in an amplifier exhibiting negative internal feedback (typically a grounded emitter circuit with an unbypassed emitter load). The impedance shown as Z_V in the supply lead is unique and has a major impact on the dark adaptation characteristic of the visual system.

The distribution amplifier is not called out specifically in Tranchina & Peskin. It is basically a unity gain inverting amplifier. There are two sources of delay in this figure, the intrinsic delay of the P/D mechanism and the transport delay associated with the distance between the distribution amplifier and the pedicle, the length of the photoreceptor axon.

[16]Tranchina, D. Peeskin, C. (1988) Op. Cit. figure 4

Tranchina adopted the analytical tools developed by the Baylor school to describe the waveforms of the P/D Equation, to the analysis of the waveforms measured at the output of the horizontal cells of the retina. This led him to define multistage filters with high values of n. While a 1990 paper with Purpura, et. al. presented the calculated responses of the monkey visual system at the location of the LGN based on these filters, they did not present any data points on the same graphs to demonstrate the appropriateness of their analysis[17]. To realize delays to peak response of 50 to over 150 ms as observed experimentally, typically 25-34 stages were required in the filters. Their analytical work was also limited by the range of signaling theory that they employed (See **Section 16.1.3**).

As will be developed further in **Section 16.4**, the use of the complete P/D Equation to define the intrinsic delay due to transduction and the calculation of the transport delay associated with the axons and dendrites involved lead to a significantly different characterization of the overall situation.

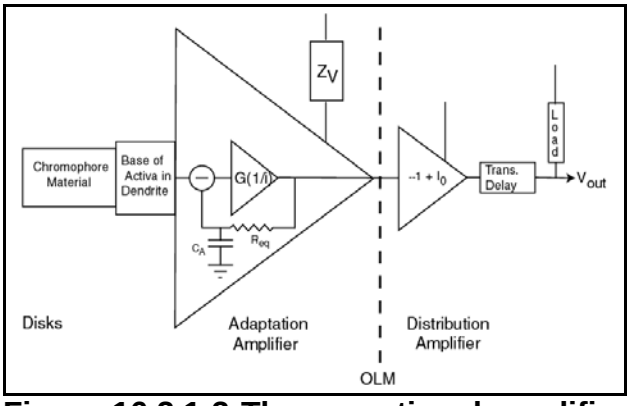


Figure 16.2.1-3 The operational amplifier architecture of the photoreceptor cell. It combines the signal transduction of stage 1 with the initial amplification and signal translation of stage 2. The transduction stage incorporates an intrinsic delay. The translation stage (and subsequent neural circuits incorporate a transport delay. Conversion from an excited electron to a free electron occurs in the base region of the Activa within each dendrite. The feedback loop is internal, there is no physical subtraction mechanism. The distribution amplifier provides a current gain of -1 (inversion) plus a bias current, C. The voltage at the pedicle is formed by the above current passing through a diode load.

16.2.1.7 The formation of the saliency map

Each of the signaling channels of vision contribute to the formation of the visual portion of the saliency map within the brain. Our understanding of this process remains conceptual at this time and will not be addressed here. See **Chapter 15**.

16.2.2 Old and new equations of vision

16.2.2.1 Previous obsolete conceptual equations

As discussed in **Section 17.3.3**, the field of colorimetry has evolved from the conceptual premise that human color vision could be described by the linear summation of three spectral components. The premise has been frequently presented mathematically as:

$$Q = xR + yG + zB \qquad \text{Eq. 16.2.2-1}$$

where Q is the "total color," R G and B are the signal components related to the red, green and blue channels of trichromatic vision and x, y & z are in percent. This formulation has been historically called the Equation of Color Vision, the Equation of Color, or the Color Equation. It will be used here as the simplest conceptual way to express the external stimulus to the eye, as well as the net signal in the luminance channel of vision when expressed in linear amplitude space. Its equivalent in logarithmic amplitude space will be R(λ,F). When required, the sum of the spectral components will be labeled Q(λ,F)

---

[17]Purpura, K. Tranchina, D. Kaplan, E. & Shapley, R. (1990) Light adaptation in the primate retina: analysis of changes in gain and dynamics of monkey retinal ganglion cells. *Vis. Neurosci.* Vol. 4, pp 75-93

in this work to emphasize its dependence on wavelength and the intensity of the stimulus.

The above equation has appeared in many expanded forms. Usually, the goal is to define the contributions to each of the terms R, G and B. One form applying to a specific element of the retina at position x is[18]:

The symbol § will be used in the following equations to indicate a summation because of an incompatibility between WordPerfect and Adobe Acrobat.

$$\rho^x_k = \S\ R_k(\lambda_n)E(\lambda_n)\,S^x(\lambda_n) \qquad \text{Eq. 16.2.2-2}$$

where R is the spectral sensitivity of the k-th photoreceptor type, E is the intensity of the source irradiance expressed in quantal units, S is the reflectance of the surface of the object at position x and n is the index of the wavelength λ. The terms are summed over the interval of n for each of the k types of spectrally sensitive photoreceptors. The result, ρ^x_k, describes the "quantal catch" in the k-th spectral channel at the location, x.

As discussed in detail in **Section 17.3.3**, neither of these conceptual forms has ever been related to a specific functional model of, or to a specific perceptual plane in the architecture of, the visual system. Occasionally, the second expression has been associated with the output of the photoreceptor cells[19]. It will be shown this association is inappropriate. Unfortunately, equations of these forms are at best of pedagogical interest. They do not provide an adequate foundation for describing either the luminance or chromatic aspects of vision. The appropriate equations are developed in the remainder of this Chapter.

16.2.2.2 Equivalent steady state functional equations

16.2.2.2.1 The equations of stage 1, the signal detection stage

The equation of Wandell can be made more useful in a few steps. First, the equation specifies a bin spacing of the wavelength samples but does not introduce the bin width, Δλ. Second, it does not include a term to account for the absorption of the physiological optics. Finally, it should be reinterpreted to include a ultraviolet channel. The resulting equation is shown below.

$$\rho^x_k = \S\ E(\lambda_n)\,S^x(\lambda_n)O_k(\lambda_n)\,R_k(\lambda_n)\,\Delta\lambda \qquad \text{Eq. 16.2.2-3}$$

where $O_k(\lambda)$ describes the absorption of the optics and k describes the four channels of the visual system, U (ultraviolet), S, M and L. The summation only refers to wavelength. The values of ρ^x_k are not summed.

In this form, the values of ρ^x_k are on a per unit area basis. To move toward a more useful form, the appropriate equivalent area for each channel is needed. This will be described on a per receptor basis at this time. It is typically the absorption cross-section of the outer segment. By taking the limit as Δλ becomes small, the equation can be converted to an integral form of more general use.

$$\rho^x_k = \S\ \text{Integral}\ E(\lambda_n)\,S^x(\lambda_n)O_k(\lambda_n)\,R_k(\lambda_n)\,d\lambda \qquad \text{Eq. 16.2.2-4}$$

The equation in this form describes the "quantal catch" per unit time of each photoreceptor channel related to the transduction function (not including the translation function) of this work. When transitioning to a current related to the free electrons within the neural circuit, the unique characteristic of the long

---

[18]Wandell, B. (1989) Color constancy and the natural image, *Physica Scripta*, vol. 39, pp 187-192

[19]Brainard, D. & Wandell, B. (1986) Analysis of the retinex theory of color vision, *J. Opt. Soc. Am. A.* vol. 3, no. 10, pp 1651-1661

wavelength channel must be introduced (see Section xxx). This is most easily done by separating the equation into two expressions and defining the "electron catch," σ^x_k and a conversion efficiency term, K.

$$\sigma^x_k = K\rho^x_k = \S \text{ Integral } E(\lambda_n)\, S^x(\lambda_n) O_k(\lambda_n) R_k(\lambda_n)\, d\lambda \qquad \text{Eq. 16.2.2-5}$$

where the equation describes the free electrons generated in only the k-th spectral channel designated by U–, S– and M–.

$$\sigma^x_L = K\rho^x_L = [\text{Integral } E(\lambda_n)\, S^x(\lambda_n) O(\lambda_n) R_L(\lambda_n)\, d\lambda]^2 \qquad \text{Eq. 16.2.2-6}$$

is the equation for the free electrons generated in the long wavelength channel, L–.

Neumeyer & Arnold have confirmed that the tetrachromatic eye of the goldfish exhibits the same performance as suggested by the above equation for the long wavelength trichromat. They found that the goldfish losses sensitivity in the long wavelength region under conditions of reduced overall light (1.5 lx instead of 25 lx white room illumination)[20]. As a result of the loss of long wavelength sensitivity as a function of light level, it would be expected to lose one dimension of its total chromatic capability as well.

16.2.2.2.2 The equations of adaptation within stage 2

After forming these equations, it is appropriate to move on to the output of the adaptation amplifier circuit. Within its operating range, the adaptation amplifier circuit introduces two distinct factors into the signal path of each spectral channel. First, it introduces a gain that is inversely proportional to the applied signal amplitude. This represents the adaptation function. It also introduces a typically large amount of negative feedback into each channel of the system. This feedback passes through an RC filter that limits its impact to slow changes in the signals applied to the amplifier. As a result of these actions, the above equations become:

$$\tau^x_k(t) = G_k\sigma^x_k(t) - G_k \text{ Integral } \sigma^x_k(t)\, dt \qquad \text{Eq. 16.2.2-7}$$

$$\tau^x_L(t) = G_k\sigma^x_L(t) - G_k\text{Integral } \sigma^x_L(t)\, dt \qquad \text{Eq. 16.2.2-8}$$

where G_k and G_L are typically inversely proportional to σ^x_k and σ^x_L over the photopic region of vision and the integrals are evaluated over the time constant, T, of the feedback network.. The introduction of the variation with respect to time, t, will be addressed below. The internal operation of the feedback circuit tends to cancel the exponent associated with the σ^x_L term (See **Chapter 12)**.

The critical feature of both these equations is the fact that the output of both channels approaches zero for slow changes in the stimuli, σ^x_k or σ^x_L. As a result, the nominal biological visual system is blind to slow changes in the product of $E(\lambda)$ and $S^x(\lambda)$ (See **Section xxx**). The visual systems found in *Arthropoda* and most members of *Mollusca* are highly responsive only to changes in one or the other of these two quantities.

16.2.2.3 The temporally simplest functional equations

The eyes of *Chordata* and some members of *Mollusca* have incorporated an additional mechanism, tremor, to overcome this limitation. By introducing this mechanism, the fundamental expression for the product of $E(\lambda)$ and $S^x(\lambda)$ is modified to

$$[E(\lambda)S(\lambda)](t) = \{ [E(\lambda)S^x(\lambda)] - [E(\lambda)S^{x-1}(\lambda)]\} / \Delta x/V. \qquad \text{Eq. 16.2.2-9}$$

where Δx is the angular distance between adjacent scene elements and V is the equivalent angular velocity of the nominal tremor in object space.

This mechanism essentially transforms the spatial frequency of every component in object space from zero to a nominal temporal frequency within the passband of the adaptation amplifier. **Without the presence of**

---

[20]Neumeyer, C. & Arnold, K. (1989) Tetrachromatic colour vision in goldfish and turtle. **XXX** pg. 623

tremor, the chordate eye is blind. With tremor, the chordate visual system becomes sensitive to changes in the texture of a scene as well as to rapid changes in overall illumination or reflection. Such a system is best described as a change detector. Note that it remains insensitive to areas of uniform reflectance within a scene. **In this sense, the chordate visual system is an edge detector <u>only</u>**.

By substituting this equation in **Eq. 16.2.2-5** & **-6**, Eq. **16.2.2-7** & **-8** become

$$\tau^x_k(t) = G_k\sigma^x_k(t) + G_k\sigma^{\Delta x}_k(t)(V/\Delta x) - G_k \text{ Integral } \sigma^x_k(t)\, dt \qquad \text{Eq. 16.2.2-10}$$

$$\tau^x_L(t) = G_k\sigma^x_L(t) + G_L\sigma^{\Delta x}_L(t)(V/\Delta x) - G_k \text{ Integral } \sigma^x_L(t)\, dt \qquad \text{Eq. 16.2.2-11}$$

[xxx Δ apparently is not defined in superset codes]
where the first term on the right includes changes with time alone and the second term includes changes in time related to changes in spatial reflectance or illuminance.

It can be shown that under small signal (low contrast) condition, the integrator in the adaptation amplifier feedback loop makes G_k independent of variations in either $E(\lambda)$ or $S(\lambda)$.

16.2.3 The equations at the pedicles of the photoreceptor cells

The pedicles play a unique role in the electrophysiology of vision. They introduce an intrinsic nonlinearity into the visual process. This nonlinearity (along with the adaptation mechanism) are key to the phenomenon of color constancy.

16.2.3.1 The output currents at the pedicles

The distribution amplifiers within the photoreceptor cells perform a simple signal inversion in their role as part of the differential pair circuit they share with the adaptation amplifiers (see **Section xxx**). In this role, they essentially subtract the output of the adaptation amplifiers from a constant value. The current gain of this process is nominally 1.00.

$$\nu^x_k(t) = 1 - G_k\tau_k(t) \qquad \text{Eq. 16.2.2-12}$$

$$\nu^x_L(t) = 1 - G_L\,\tau_L(t) \qquad \text{Eq. 16.2.2-13}$$

16.2.3.2 The output voltages at the pedicles

The more important function at the output of the distribution amplifier is the conversion of the signal at the pedicle from a current to a voltage. This process aids in the signal manipulation to follow. It is the voltages at the pedicle that are the signals used by the signal processing of stage 2. The process involves an electrical load impedance in the form of a diode. The process causes a change in the basic form of the equations. They can be written as the logarithm of the above forms.

$$\omega^x_k(t) = C_k \text{ Ln } \nu^x_k(t) = 0 - \text{Ln}[G_k\,\tau^x_k(t)] \qquad \text{Eq. 16.2.2-14}$$

$$\omega^x_L(t) = C_L\text{Ln } \nu^x_L(t) = 0 - \text{Ln}[G_L\,\tau^x_L(t)] \qquad \text{Eq. 16.2.2-15}$$

This logarithmic conversion is the fundamental non-linearity of the visual process. It affects all of the subsequent signal manipulations within the visual system.

This set of equations (after expanding by replacing the tau functions) are the functional equivalent of the conceptual equation (3) of Wandell given in **Eq. 16.2.2-2.** However, it is important to note that these equations do not represent the absolute steady state as suggested by Wandell's discussion. These equations are only relevant following introduction of a temporal or spatial change in the signals applied to the individual spectral channels of the photoreceptors. Such a change may be due to the normal tremor associated with the eyes of many chordates.

Except for certain DC voltage offsets and 2nd order temporal effects, this set of equations completely describe the nominal voltages at the pedicles of the photoreceptor cells following transient excitation.

Wandell proceeded to present his conceptual equations in matrix form (after caveating his material with respect to linearity) by separating the lights and surfaces components from the sensory components [his equation (4) and figure 6.5]. This is not easily done with the functional equations because of the nonlinearities associated with the adaptation process and the logarithmic transformations at the pedicles. It cannot even be done under the small signal model within the photopic region because of the high gain introduced by the adaptation process. Using factoring to create a matrix based approximation to the actual situation introduces considerable confusion with respect to the perception of metameres.

It will be shown below that equation (4) of Wandell cannot be associated with the functional properties of the pedicles of the photoreceptor cells. The linear summation of the spectral signals associated with the individual spectral photoreceptor types never occurs in the visual system.

16.2.4 The equations of stage 2, signal processing

The signal manipulation occurring subsequent to the pedicles of the photoreceptors leads to the signals used in the perception of both the brightness and chromatic properties of a scene. The combinations of these voltage signals that are important in vision are shown in **Figures 16.2.1-1** and **16.2.1-2** . The signal manipulation subsequent to the photoreceptor cells is linear in the first order. Therefore, matrix notation can be used effectively to describe the signals of stage 2 and beyond in the visual system.

16.2.4.1 The signals at the S-plane (perceptual plane) of vision

The first recognizable occurrence of the brightness and chrominance signals used in the perception of vison is at the so-called S-plane of the retina. The G'–channel signals associated with geometric perception of fine detail are not formed within the retina. These signals from the foveola are sent directly to the midbrain without any signal processing. The initial processing of the signals associated with the foveola occurs in the pretectum of the midbrain.

As a result of the processing within the retina, multiple parallel signal paths are created between the retina and the brain. The paths associated with chrominance (and many other features except brightness) are formed in differential circuits. The resultant signals are frequently labeled opponent signals although a more descriptive term would be differential signals.

The brightness signal, R, is formed by the summation of voltage signals from the pedicles of the photoreceptor cells.

Both the brightness and other signals present at the S-plane are analog signals. Any conceptual designation related to inhibition should be avoided when speaking of these signals.

In general, the signal processing related directly to the brightness and chrominance channel signals do not involve any significant differential time delay within the inner nuclear layer of the retina. This may not be true with respect to other signals processed by the second lateral processing matrix.

16.2.4.2 The transfer of signals between the pedicles and the S-plane

There is little experimental data available on the transfer of the signals from the photoreceptor cells to the output of the horizontal and bipolar cells at the S-plane. The model tells us the signal transfer process is similar to that used in TTL (transistor-transistor logic) electronic circuits. The signal from a pedicle is passed through a first transistor (Activa) to the input structure of a second transistor (Activa) within either a horizontal or bipolar neuron. The first Activa is designated a synapse morphologically. The synapse is an Activa acting as a diode. The input structure of the second Activa can also be described in terms of a diode. However, in the case of the inverting input to the Activa of the horizontal cell, the diode is generally shunted by a second impedance. This impedance is also typically a diode. **Figure 16.2.1-4** shows the typical circuit for a bipolar cell in the R–channel. The components labeled Z_x are the synapses associated with the individual spectral channels and the component labeled Z_{eq} is the input impedance of the Activa within the cell. Since current only flows in one direction through these devices, each serial connection of a Z_x and the Z_{eq} effectively forms an isolated voltage divider. The precise performance of these voltage dividers depends on the size of the synapses and the Activa within the bipolar cell. It also depends on the current through each of them. The voltage at the input to the second Activa compared to that at the pedicles can be described by m_{XU}, m_{XS}, m_{XM} and m_{XL} where the X describes which signaling channel the voltage

refers to. In general, the non-inverting voltage gain of the second Activa cannot exceed unity.

In the case of the inverting input, the gain of the second Activa can exceed unity. However, the voltage divider associated with the inverting input generally provides a smaller potential to the Activa terminal. The available data suggests the net gain of the inverting signal path is unity by design. To a first approximation, the overall gain associated with the inverting input terminal will be considered unity. This condition simplifies the description of the overall set of signals at the S-plane.

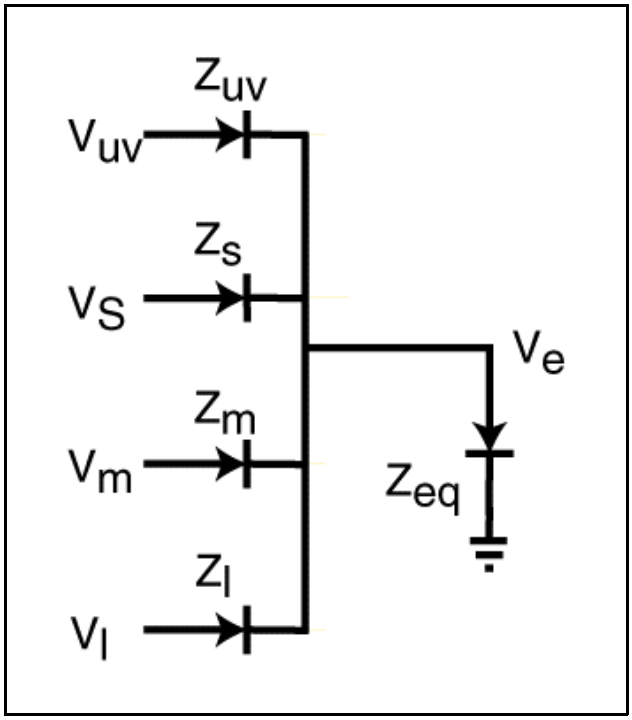


Figure 16.2.1-4 The summing circuit at the output terminals of four photoreceptor cells. It can be described as a "soft" analog "OR" circuit. See text for details.

16.2.4.3 The signals of the S-plane in matrix form

A fundamental principle of biological vision is that the information collected by the photosensitive photoreceptors is processed into a series of parameters that are treated as orthogonal members of a single vector space. Based on this principle, the fundamental equations can be written as a single multidimensional matrix with the products treated as individual parameters in a multi-dimensional vector space.

The most important of these are the Luminance equation and the appropriate set of Chrominance equations. However, there are also Polarization (N–channel) and Appearance (G–channel) equations that must be described for some animals. The matrix can be expanded to incorporate these channels when discussing the performance of specific species. A subset of the Appearance equations is particularly important in Hominoidea. The equations related to the G'–channel account for the unique spatio-analytical capability of this family. This is the ability to examine and interpret fine detail and symbols–in short, to read. The equations of the G'–channel introduce a great deal of complexity into the above matrix. They will be discussed separately.

Based on the above discussions, the signals at the S-plane of a tetrachromat can be provided in a matrix form where *O, P Q* and *R* are the voltages in the respective signaling paths, O–, P–, Q– and R– at the S-plane.

$$\begin{bmatrix} O \\ P \\ Q \\ R \end{bmatrix} = \begin{bmatrix} \omega_U \\ \omega_S \\ \omega_M \\ \omega_L \end{bmatrix} \cdot \begin{bmatrix} m_{OU} & -m_{OS} & 0 & 0 \\ 0 & m_{PS} & -m_{PM} & 0 \\ 0 & 0 & -m_{QM} & m_{QL} \\ m_{RU} & m_{RS} & m_{RM} & m_{RL} \end{bmatrix}$$

Matrix form of The Vision Equations for a tetrachromat Eq. 16.2.4-1

The above matrix is the correct form to express **The Equations of Vision**. It replaces the single conceptual equation of Eq. 16.2.2-1. **It correctly describes both the perception of brightness and the complete perception of color in the biological visual system**. The superscripts, x, related to position and the dependence on time have been omitted from the two left-most matrices for convenience. The expression can be evaluated at any point in object space where a change has occurred in illumination, reflectivity or with spatial position. The coefficients in the right-most matrix used to form one signal (brightness) need not have the same value as the equivalent coefficient used to form a different channel (chrominance). More specifically, there is no direct (or fixed) relationship between the coefficients associated with the brightness equation and those associated with the chrominance equations. This matrix will be justified in the following sections.

By expanding the above matrix, individual equations can be obtained for the perceived brightness, R, of a scene element as well as the equations for the chromatic parameters, O–, P– & Q–, of a three dimensional color space. This color space is similar to a two-dimensional Hering color space. See **Section 17.3.3**.

The parameters in this comprehensive matrix need not be constant in time or with respect to other parameters. Thus, it is appropriate to write the expansion of this matrix in integral, temporal, steady- state and other forms as the experimental and/or pedagogical situation requires. In the following matrices, all coefficients are positive. The negative signs within the matrices are indicative of the subtraction process.

16.2.3.4.1 A simplified (degenerate) matrix

Why has man previously discussed his own visual system as trichromatic? The answer becomes clear if the equation for ω_U is expanded back to the level of Eq. 16.2.2-4. It is seen that ω_U is dependent upon not only the sensitivity of the photoreceptors to ultraviolet light but also the presence of ultraviolet photons in the illuminant, a significant variation in the reflectivity of the scene in that region of the spectrum and reasonable transmission by the physiological optics. Each of the later parameters can negatively impact the performance of the biological visual system If the illumination is deficient in ultraviolet photons, the system will appear to be trichromatic. This is the norm when testing within the laboratory (or being inside

any building without a window with a northern exposure). Even with a northern exposure, the glass o f the window and the lighting system is a poor transmitter of ultraviolet photons. Because of the actinic properties of ultraviolet photons, the variation in reflectivity of objects in the ultraviolet is usually quite low (except for newly prepared surfaces–such as fresh flowers). Finally, the physiological optics absorbs significantly in the ultraviolet. The absorption is proportional to the thickness of the physiological optics. Each of these parameters contributes to the casual assumption that humans are insensitive in the ultraviolet (particularly in the region between 395 nm and 437 nm). However, it is the external parameters that most limit the performance of humans in the above region.

As a result of the normal combination of external parameters, the term ω_U that has a very low value in spite of the very high sensitivity of the ultraviolet spectral channel of the human retina. This high sensitivity will be addressed below. Because ω_U approaches zero in the typical (laboratory) environment, the above matrix equation can be simplified by eliminating the left column in the 4 x 4 matrix. While this simplification does not completely eliminate signals from the O–channel, it makes it very difficult to evaluate the perception of any signal in this channel.

16.3 The functional equations of each signaling channel

Based on the continuity of the model of this work, similar forms of the fundamental equations can be written that apply at different morphological and electro-physiological locations within the visual system. If written at the output of the stellate cells of the *neo-* (or occipital) cortex they provide an initial insight into the memory function of the animal. At this location, the equations represent a portion of the saliency space (the memory map) of the animal.

Except in the area of memory, **the visual system is a change detection system**. Therefore, the most important matrix forms are those expressed with respect to time or temporal frequency. However, the system has traditionally been looked at as a static imaging device (which it is not). Until the field as a whole makes the transition to the fact that the visual system is a change detection system, it is easiest to discuss and teach based on the steady-state forms of the equations. There is a serious problem at the research level however. The visual system is a change detector, not an imager. Some change with respect to time must be introduced into every signal falling on the outer segment of a photoreceptor cell or no signal will be generated at the output of that cell. This fact requires careful attention to detail in the next section.

While writing the fundamental equations in matrix form as they apply to the saliency space of the animal is feasible, the complexity of the individual terms in the matrices are quite complex (primarily due to the asymmetries introduced by the projection neurons of Stage 3). Writing the equations at the input to the ganglion cells of Stage 3, that are located at the morphological output of the retina, will generally provide much of the same information in a simpler form. This approach will be followed here, except when discussing certain time dependent idiosyncracies of the visual system, such as flicker effects based on motion. To be more specific, the equations will be written at the S-plane of the retina, this plane is represented by the pedicles of the horizontal, bipolar and amercine cells generally associated with the inner plexiform layer of the retina.

In the following material, a named function will always relate to a function of wavelength. Unfortunately, the term function will frequently be used without a name describing its role in vision. This is necessary because of its ubiquitous use in mathematics. An equation will always relate to the result of an integral with respect to wavelength.

The remainder of this section will be subdivided into:

16.3.1 General background material,

16.3.2 the Luminance Equations and Functions of a tetrachromat,

16.3.3 the reduction of the above equations applicable to a blocked tetrachromat (trichromat),

16.3.4 the Chrominance Equations and Functions,

16.3.5 the Appearance Functions and

16.3.6 the Temporal Equations of Vision.

In **Sections 16.3.2** through **16.3.5**, the role of changes in the scene related to time (or tremor) will not be stressed (for the sake of simplicity). However, it must be recalled that the visual system is blind in the absence of such short term changes. These forms will all include integrals with respect to wavelength. **Section 16.3.6** will re-introduce the temporal forms of the underlying equations needed to address the transient performance of the visual system. **Section 16.3.7** will develop the mechanically dynamic aspects of the visual system.

To minimize complexity, the equations of **Sections 16.3.1** through **16.3.xxx** will not include the absorption due to the physiological optics of the eyes. They will also avoid introducing the variation in gain associated with the adaptation amplifiers. This variation (key to color constancy) will be introduced in **Section 16.xxx.xxx**.

16.3.1 Practical and experimental limitations

Experimentalists have long overlooked the critical nature of tremor in the performance of the visual systems of biology. Most workers in the field overlook the fact that Tomita used a chopper in his optical path to modulate the light applied to the retinas of his fish. This methodology insured that his results reflected the performance of the visual system whether any tremor was present or not. If he had not chopped the light from his source, his data may have shown more variation, with respect to the temporal aspects of his instrumentation and the animals response system prior to the S-plane, than he was prepared to explain.

As suggested in **Section 16.2.4.2**, there is some cross-coupling in the summation process within the luminance channel since the impedance of the element, Z_{eq}, is a function of the current through it. This can cause some difficulties when measuring the luminance sensitivity of the eye under conditions of chromatic adaptation. However, this cross-coupling is insignificant when narrow band spectral illumination is used to measure the spectral sensitivity function (also known as the Luminous Efficiency Function). The problem of cross-coupling is smaller within the chrominance and other differential channels.

16.3.1.1 Subdividing the model

Looking at **Figures 16.2.1-1 & 16.2.1-2**, the architecture is seen to change at the location of the pedicles. This is the physiological justification for the zone theory of color vision. The individual signaling paths subsequent to the pedicles can be subdivided into a series of cascaded terms just as the spectral channels prior to the pedicles were. Fortunately, the portion of the overall system within Stage 2, signal processing, is relatively simple. There is little or no cross-coupling of signals. Stage 3, signal projection, is much more complex with regard to potential cross-coupling.

16.3.1.2 Eliminating marginal channels

While the complete equations of the tetrachromatic eye can always be applied to a biological vision of a given species, it is frequently possible to eliminate one or more of the spectral or signaling channels of the general model. This is commonly done with respect to humans for two reasons. The ultraviolet sensitivity of the system following the lens has not been appreciated and the light sources usually used do not provide significant light in the 400-437 nm region. The absorption of the lens in the ultraviolet is so great that the gain of the adaptation amplifier cannot compensate for it. For these reasons, the ultraviolet capability of the human visual system has generally been ignored. When performing precise evaluations, the spectral absorption by the U–spectral channel in the region between 400 and 437 nm must be considered. It affects the signal level in the O–channel. This signal is important in quantifying the response of the human system in the purple region defined in this work.

A less significant limitation of the physiological optics of humans is better documented but seldom treated properly. This is the absorption of the macular in the short wavelength region of the spectrum. This absorption is normally compensated by the adaptation amplifier of the S–channel. The result is only a narrowing of the overall photopic operating range.

The geometric channels of the human visual system (other than the G'–channel related to the foveola) has also been largely ignored. This has been due to the limited performance of the human visual system in this regime.

16.3.1.3 Quantifying the laboratory stimulus

In the general case, *E* is an irradiance defined in terms of a number of photons per second per unit area per unit wavelength on the focal plane surface of a "receiver". As such, it can extend over a wide spectral range and have any variation in intensity with wavelength desired. If one is speaking of a non-focal location (such as a pupil), the expression becomes more complex. It must include a description of the cone angle related to the incident photon flux. This freedom in the value of *E* defined above complicates the discussion significantly. It is useful to consider limiting the variations in E to just a few mathematical forms.

+ The most important form is to set it equal to a constant intensity per unit wavelength. At this value of *E*, each of the photoreceptors of vision receive an equal number of photons per second per unit wavelength. Such a constant value is achievable over the entire spectrum of interest in vision research. The discussion of the quality of this value is presented in **Section 2.1.1**. The preferred color temperature is higher than commonly used in prior research. The optimum choice is between 7053 and 8000 Kelvin. If a lower temperature is used, a correction factor must be included in the analysis.

+ Since much of the research appearing in the literature has employed an incandescent source of undetermined color temperature, but probably near 2856 Kelvin, it is appropriate to consider this color temperature for *E* also. A source corresponding to Planck's Law is known as a *full radiator* in the parlance of the C.I.E. A *full radiator* at absolute temperature of 2856 Kelvin is defined as C.I.E. Standard Illuminant "A"

+ To be compatible with much prior research, it is advisable to evaluate the equations of vision for *E* equivalent to Illuminant "C." The C.I.E. has defined Standard Illuminant "C," (1931) in terms of a correlated color temperature they specify as 6774 Kelvin. This illuminant was meant to characterize daylight in the absence of direct solar illumination. While such a source or direct solar illumination may meet an equal energy per unit wavelength criteria, they are significantly lacking when based on an equal flux per unit wavelength criteria. They are both lacking in the blue and ultraviolet regions.

> It is important to note the C.I.E. distinguishes between sources and illuminants. Whereas a source is a specific physical radiator, an illuminant refers to a specific spectral radiant power distribution, that may not be physically realizable in the laboratory.
>
> It is also important to note that the C.I.E. defines a correlated color temperature that is not defined by the spectral radiation from a true black body. This subject is discussed in detail in Wyszecki & Stiles[21]. A correlated color temperature is defined in terms of the C.I.E. Chromaticity diagram coordinates. It is unfortunate that the C.I.E. has begun defining standard illuminants in this way using a Diagram that suffers from a lack of conformality and is itself defined using imaginary visual stimuli.
>
> Note should be made of the intent of the C.I.E. to declare Illuminant"C," and also "B," obsolete because of their serious deficiencies in the ultraviolet region. This region is playing an ever more important role in the commercial application of modern colorimetry. It is critical to the standardization of fluorescent materials.
>
> Because of the difficulty in creating a source or an illuminant conforming to the specifically defined spectral radiant power distribution of the newer "D" series of illuminants, the C.I.E. has not recommended any Standard Source that can be used to produce these illuminants. Instead, it has prepared guidelines for evaluating a potential source referred to as *daylight simulators of colorimetry* (1980).
>
> In general, any source used in evaluating the S-channel of vision must not use a normal glass envelope. A fused quartz envelope is required to achieve the desired short wavelength transmission.
>
> Because of the above notes, it is incumbent on any laboratory investigator to carefully define the source, or illuminant, used in their experiments. This is especially true if the

[21]Wyszwcki, G. & Stiles, W. (1982) Color Science 2nd. Ed. NY: John Wiley & Sons. Chapter 1 and pp. 143-155

performance of the S-channel is a factor in the investigation.

Wyszecki & Stiles, fig. 3(3.3.4), provide a diagram of the locus of a Planckian radiator and a C.I.E. daylight locus "D" superimposed on a Chromaticity Diagram (1931).

16.3.2 The luminance equation of the tetrachromatic and aphakic eye

Looking at **Figures 16.2.1-1 & 16.2.1-2**, the signal in the luminance, R–, channel can be defined as the sum of the signals received from all of the functioning spectral channels of the subject. Using the notation developed in the previous section, the equation for this signal can be written as,

$$R(t) = m_{RU}(\omega_U) + m_{RS}(\omega_S) + m_{RU}(\omega_M) + m_{RU}(\omega_L) \qquad \text{Eq. 16.3.2-1}$$

with the t shown explicitly to remind the reader of its importance relative to the time constant of the adaptation amplifiers.

This equation is linear with respect to the potentials of the photoreceptor pedicles. It is the same as the equation for the R–channel in the matrix of **Section 16.2.4**. The equation can be expanded by substituting for the omega terms in accordance with **Section 16.2**. However, the equation quickly becomes nonlinear with respect to a variety of parameters including the input stimulus. It can be written as,

$$R(t) = m_{RU}(C_U \operatorname{Ln} \nu_U) + m_{RS}(C_S \operatorname{Ln} \nu_S) + m_{RM}(C_M \operatorname{Ln} \nu_M) + m_{RL}(C_L \operatorname{Ln} \nu_L) \qquad \text{Eq. 16.3.2-2}$$

Using the more compact form of the previous section, and omitting a negative sign in each term,

$$R(t) = \S\, m_{Rk}(C_k \operatorname{Ln} \nu_k) = \S\, m_{Rk}(C_k \operatorname{Ln} G_k \tau_k) \qquad \text{Eq. 16.3.2-3}$$

over k = U, S, M & L in turn. Alternately,

$$R(t) = \S\, m_{Rk}(C_k \operatorname{Ln} G_k \sigma_k(t))$$

The total function becomes

$$R(t) = \S\, m_{Rk}(C_k \operatorname{Ln} G_k \text{ Integral } E(\lambda)\, S(\lambda) O(\lambda)\, R_k(\lambda)\, d\lambda) + m_{RL}(C_L \operatorname{Ln} G_L [\text{Integral } E(\lambda)\, S(\lambda) O(\lambda)\, R_L(\lambda)\, d\lambda]^2) \qquad \text{Eq. 16.3.2-4}$$

for k = U, S & M.

The above equation is the <u>complete</u> equation for the perceived illuminance (brightness) of any biological vision system. It applies to all levels of intensity and spectral distribution that can be presented to the pupil of the visual system.

> It is important to note that the sum of the logarithms is not equal to the logarithm of the sum. This is a crucial postulate of mathematics and a crucial factor in the perception of a scene. It is both the cause and explanation for the Bezold-Brucke and the Purkinje phenomena.

Note that because of the overlap in the spectral performance of the various chromophores, the terms within the summation sign are not independent. This overlap is the source of the Brezold-Brucke Effect. The impact of this effect is particularly noticeable when the spectral sensitivity function is measured using narrowband irradiance. Note also the greater sensitivity to change in the L–channel. This is the explanation for the perceived hyperactivity of flames rising from an uncontrolled fire. It is also related to the choice of a red color for stop lights.

The overall function is defined in terms of the voltage at the pedicle of the bipolar cell associated with a scene element in object space at least as large as the arborization of the bipolar cell. In the general case, <u>*E* is the irradiance incident (in photons per second per unit area per unit wavelength) on the scene element in object space and S is the ratio of the incident irradiance to the reflected irradiance reaching the aperture of</u>

the physiological optics of the eye. S is generally a very small fraction. O represents the transmission of the physiological optics as a function of wavelength. R_k remains the absorption coefficient of the outer segments of a particular chromatic channel.

16.3.2.1 The luminance equation within the photopic region

Within the photopic region, the above equation can be simplified. The adaptation mechanism is designed to maintain the product of the gain and the average of the integral of the product of E, S, O and R_k in each spectral channel equal to 1.00. As a result, using Equation 16.2.2-9, R(t) can be written as;

$$R(t) = \S\, m_{Rk}\, C_k \,\mathrm{Ln}\, G_k \,\mathrm{modulation}(t) + m_{RL}(2C_L \,\mathrm{Ln}\, G_L \,\mathrm{modulation}(t) \qquad \text{Eq. 16.3.2-5}$$

for k = U, S & M and

where the modulation is the (absolute) variation in the above product around the mean value of the product. The numeric 2 is introduced due to differentiation of a squared term. This result is completely compatible with the concept of brightness constancy. The product of the gain and the modulation will remain a constant within the photopic range. No matter what the variations in long term variations in the irradiance, reflectance, or optical absorption, they will be compensated for within the photopic region. However an enhanced dynamic response will occur in the L–channel. A similar result will be developed below with respect to color constancy.

16.3.2.2 The spectral sensitivity of the aphakic human eye

Griswold & Stark have recently provided the luminous efficiency function for the aphakic human eye. This term describes an eye in which the lens has been surgically removed. The results are shown in [**Figure 17.2.1-1**] and are compelling. The spectral sensitivity of the human retina extends to at least 315 nm in the ultraviolet. It also exhibits a high sensitivity in the region of 347 nm as expected by the presence of the ultraviolet chromophore, Rhodonine(11). The data demonstrates that the human retina remains tetrachromatic even into advanced age.

Under the intensity conditions employed by Griswold & Stark, the nominal coefficients for the aphakic human luminance equation are , C_{RU}' = 0.3, C_{RS}' = 0.1, C_{RM}' = 1.000, and C_{RL}' = 0.1 where the prime indicates the product of all coefficients to the left of the logarithmic term in the above equations.

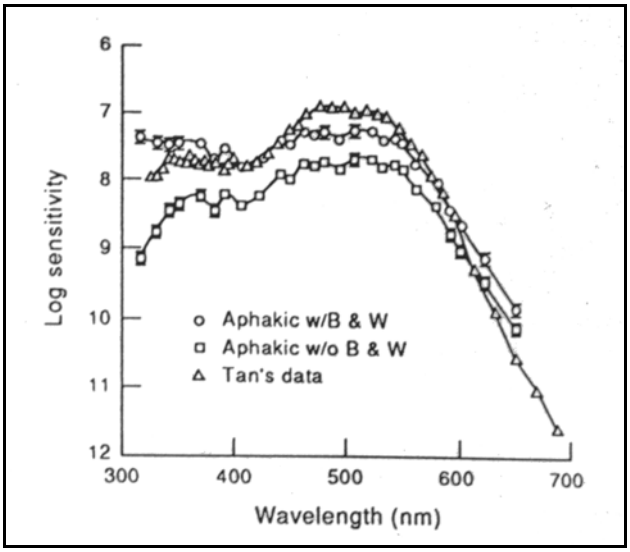


Figure 16.3.2-1 The luminous efficiency function of the aphakic human eye. From Griswold & Stark, 1992.

16.3.3 The luminance equation for the blocked tetrachromat (including humans)

[xxx edit]
The visual system of humans and other large chordates can be considered to be degenerate. The overall system was designed with a substantial capability throughout the spectral range from the ultraviolet to the near infrared. However, in order to achieve a large physical size, the lens, and to a lesser extent the cornea were scaled up. This scaling introduced significant absorption in the ultraviolet region of the spectrum, i. e., the coefficient C_U' in equation 16.3.1-5 is very small. As a result, the human visual system exhibits an ultraviolet sensitivity less than 0.1% of the peak sensitivity in the mid wavelength region. As a result of this evolution, the human eyes are considered to be trichromatic. To differentiate them from another conceptual class of trichromats, they are best described as long wavelength trichromats.

The above performance of the aphakic human demonstrates the architecture of the human eye is clearly tetrachromatic. However, the conventional wisdom is that it is trichromatic. This is primarily due to the inadequacy of the available instrumentation in the 1800's and first half of the 1900's. Since that time, more and more evidence has accumulated showing that the performance is limited mainly by the available light and the use of soda glass in light bulbs and readily available optics. When these limitations are avoided, the human, and other large chordates, can be described as blocked tetrachromats. Normally, the human is described as a trichromat due to these external limitations.

16.3.3.1 Losses in the physical optics of the eye

Wyszecki & Stiles provided an overview of the losses due to the physical optics of the human eye[22]. They stress the difficulty of making valid measurements of these parameters and only describe a verbal model of the situation. They did offer a compromise absorption spectrum for the lens of the eye, based entirely on empirical data. Norren & Vos provide a more focused and specific review.[23] They address many of the different reports and approaches to measuring these properties. Highlights of their Table 1 include the variety of techniques used and spectral ranges considered, the limited number of subjects used in each investigation and the lack of any specification on the quality of the spectral filters used in the tests. Their discussion also addresses the variation in performance with the size and position of the stimulus within the pupil of the optical system.

Norren & Vos confuse the results associated with an average observer and the defined properties of the "standard observer" of the C.I.E. They suggest that the absorption of most elements of stage 0 is relatively low except for the lens. However, they do not quantify their position. Their work does highlight one important fact. Several of the investigators struggled to define the absorption of the physiological optics in terms of a simple but poorly defined concept. Norren & Vos define an ocular density curve derived from the C.I.E. defined (smoothed) scotopic luminous efficiency function and a putative absorption spectrum of rhodopsin (the isotropic absorption of a retinoid in dilute solution). The latter was based on a Dartnall nomograph that was subsequently modified to account for "self-screening." The variance in the peak absorption wavelength was also noted as a problem in this approach. It is difficult to diagram their concept and relate it to any more comprehensive model. They noted that their conceptual reasoning was the reverse of that used by Bowmaker and Weale to achieve the same objective. The resulting functions do not represent a very precise absorption spectrum for either the lens or all of the optics.

More recently, Griswold & Stark have provided a graph of the optical density of the lens using a more objective method[24]. They make no assumptions concerning the putative absorption of a specific chromophore nor employ any defined standard as a reference. They subtracted the spectrum of aphakic

---

[22]Wyszecki, G. & Stiles, W. (1982) Color Science, 2nd Ed. NY: John Wiley & Sons, pp 107-116

[23]Norren, D. & Vos, J. (1974) Spectral transmission of the human ocular media. *Vision Res*. vol. 14, pp 1237-1244

[24]Griswold, M. & Stark, W. (1992) Scotopic spectral sensitivity of phakic and aphakic observers extending into the near ultraviolet. Vision Res. vol. 32, no. 9, pp 1739-1743

human eyes from that of phakic eyes following normalization at 700 nm Their graph extends to 315 nm, has a peak optical density of 4.0 near 354 nm, and a slope between 350 and 400 nm corresponding to a Fermi-Dirac function. The limited amount of data used prevents tracing any variation with respect to the age of the sources.

The data of Griswold & Stark strongly suggests that the fundamental absorption in the 325-375 nm region is due to the same resonances found in many molecular materials including the retinoids and caretinoids. The maximum near 354 nm is very similar to that associated with the aldehyde ligand in retinal. The secondary peak is near the 325 nm peak usually associated with the alcohol ligand in retinol. These could be associated with other retinoids and xanthophylls frequently associated with the materials of the lens.

A more functional model of the physiological optics can lead to a more precise theoretical function for the absorption of the optics. The light transmitted to the retina can be described by the incident radiation applied to the cornea and subsequently passed through the pupil of the eye minus the various absorption and scattering mechanisms. The main mechanisms are:

+ reflection at the corneal outer surface due to the difference in indexes of refraction.

+ absorption by the bulk materials other than the lens.

+ absorption by the bulk material of the lens

+ ligand absorption of the aldehydes found within the structure of the lens

+ scatter by inclusion and/or voids within the structure of the lens.

The primary mechanisms are three, the absorption of the bulk material due to water, the absorption by the aldehyde ligands of the molecules forming the lens and the Rayleigh scattering within the optical path. The first factor is proportional to the size of the animal but has not been well quantified. This factor is primarily concerned with the two extreme ends of the visual spectrum. Wyszecki & Stiles provide the absolute spectral absorption for distilled water[25]. Even for 20 mm of water, the half amplitude point is at 230 nm. The inband absorption between 300 and 1000 nm can be considered constant at 88%.

The second factor is directly related to the thickness of the lens. While this factor is trivial in small animals, it is so significant that it drastically limits the ultraviolet sensitivity in large animals (weighing more than a fraction of a kilogram). The Rayleigh scattering appears to become more important with age. **Figure 16.3.3-1** provides the constants and equations used to calculate the absorption by the aldehyde and alcohol radicals of the lens. It assumes the two absorbing species exhibit a simple Fermi-Dirac absorption profile. The absorption peak for the alcohol was taken as 325 nm. The absorption peak of the aldehyde was taken as 357 nm. The Q = 15 is arbitrary. The temperature was taken as 300 Kelvin and the calculation was performed in microns to satisfy the value for K. The above values gave a good fit with the data of the next figure.

[25]Wyszecki, G. & Stiles, W. (1982) Color Science, 2nd ed. NY: John Wiley & Sons. pg 42

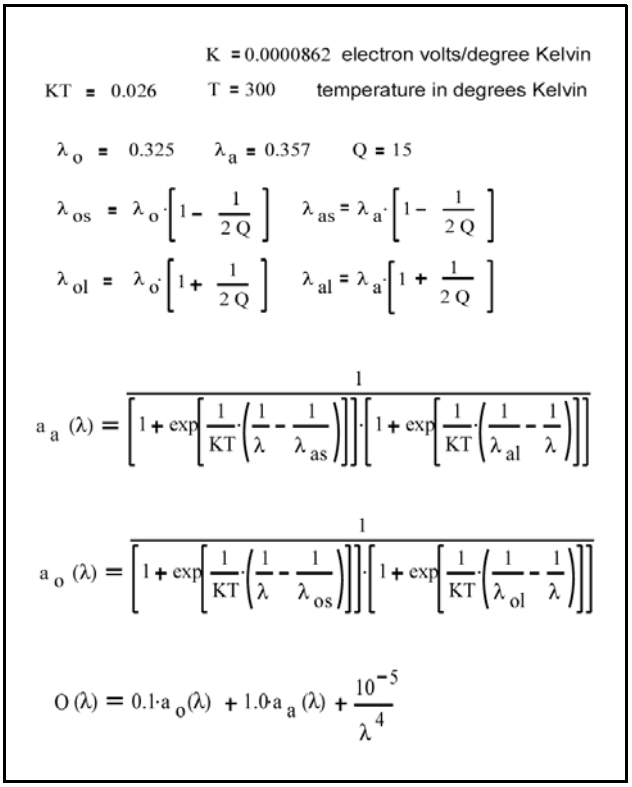


Figure 16.3.3-1 Calculation of the absorption of the physiological optics. The equation for absorption omits the small terms due to reflection at the cornea and absorption by water within the eye. A scale factor of about 5.7 is needed to fit the equation to the following graph.

Figure 16.3.3-2 compares the theoretical absorption of the physiological optics with the measured data of Griswold & Stark. The label optical density is a generic one. It includes any loss in transmission regardless of its mechanism.

The leftmost portion of the curve reflects the Fermi-Dirac absorption edge associated with a ligand of the protein material of the lens. The data from Griswold & Stark does not provide an absolute peak value nor a specific half amplitude point for the Fermi-Dirac function. However, the slope of the curve is not in question. A value of 371 nm was used for the half amplitude point of the Fermi-Dirac edge in the figure. The two sloping lines forming the rightmost portion reflects the minus fourth power relationship with wavelength of Rayleigh scattering. They differ by 10% in density of scattering particles. No statistical range was provided for the data points. However the number of subjects was quite small. It appears that additional data points between 350 and 410 nm would be useful if they were sufficiently precise statistically.

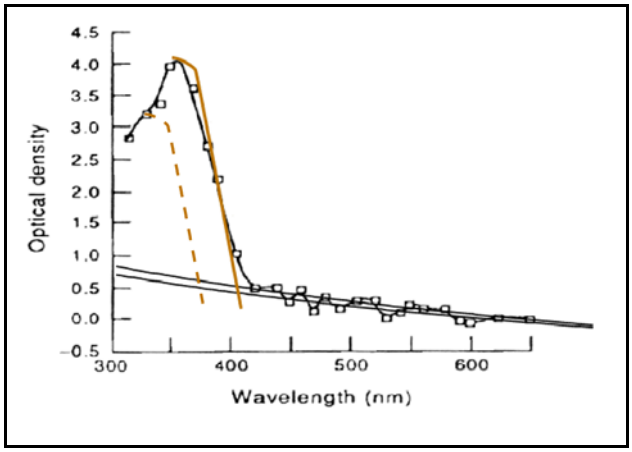


Figure 16.3.3-2 The losses associated with the physiological optics of the eye. The brown line reflects the proposed effect of ligand absorption. The dashed brown line represents the proposed ligand absorption when the aldehyde ligand is missing (see text in section 16.3.3.1.3). The pair of sloping lines represent a nominal and 110% level of Rayleigh scattering. The data points and connecting curve are from Griswold & Stark.

The above theoretical curves and the data of Griswold & Stark provide good support for the earlier compromise curve proposed by Wyszecki & Stiles. They also provide a foundation for the later explorations summarized in figure 1 of Norren & Vos. The difference between any particular investigators results and the theory can usually be assigned to the finite filter widths used in the spectral measurements or to statistical variances. Excessively broad filters result in a lowering of the slope of the measured curves relevant to the theoretical values. The variations in the steepness of the leftmost part of the curves in that figure are informative. The curve by Tan(1971) is the steepest and approaches the theoretical curve most closely.

The theoretical curves also provide a good foundation for the curves of Said & Weale (including the interpretation by Hart[26]. The reduced rates of change in slope of their curves are almost certainly due to the use of excessively broad spectral filters.

The Hart figure also shows the significant absorption related to the fovea compared to the surrounding retina. This absorption appears to be caused primarily by the changes in structure of the neural layer forming the field lens of the physiological optical system. No data could be found on the mechanism causing this absorption although the literature suggests it is bulk absorption due to the presence of different materials in this region (fewer cell nuclei and axons). See **Section 16.3.3.2.3** for an equation describing this absorption.

16.3.3.1.1 Constituents of transmission loss

The above figure is plotted with a displaced zero. Griswold & Stark provided only relative data based on

[26]Hart, W. (1992) Adler's physiology of the human eye. St. Louis, MO: Mosby Year Book,. pg 710

subtracting the measurements from phakic and aphakic eyes under the assumption that their absorption was the same at 700 nm. The difference in index of refraction between the cornea and air says that there is a 2.5% reflectance at this surface, neglecting any effect of tears. This value corresponds to an absolute loss of 0.1 units of optical density in the above figure. Although not quantified, there is also a bulk absorption due to the cornea, and the vitreous and anterior humors of the eye. As suggested by Boettner & Wolter[27], these absorption may contribute a few tenths of a unit to the total optical density of the physiological optics. The remaining loss is primarily due to the lens. This loss depends on the physical thickness of the lens. It can be quantified as about 0.7 optical density units per mm of lens thickness in the region between 315 and 375 nm. While Chen & Stark suggest that lens absorbancy does not limit the UV sensitivity in adult fish, the size of fish they used (total body length only 100 mm) would not show the effect[28].

16.3.3.1.2 Equation fitting Said & Weale data

The data for the outer lens group provided by Weale (1954) was analyzed to obtain an estimate of the effect of aging on the physiological optics[29]. If one attempts to describe this data mathematically using a first order RC type filter, the fit to the data is unacceptable. If a second order RC filter, corresponding to a high pass filter as a function of wavelength, is fitted to the data, the results appear to support the data of Weale. However, as shown above, the excessively wide spectral filters used by Weale, or the variance in his data points need to be considered. As shown above, the underlying absorption mechanism is a combination of a Fermi-Dirac absorber and Rayleigh scatterer. The upward bending of their data sets at short wavelength is caused primarily by a reduction in the quality factor, Q, of the absorption term related to the aldehyde ligand of the lens. The result is a broadening of the absorption.

By fitting the Fermi-Dirac equation to the data, one can obtain a formula that estimates the change in total absorption as a function of wavelength with a parameter of age.

16.3.3.1.3 Observed variations from the Said & Weale data in humans

A recent communication from 61 year old Robert M. (2014), responding to this work, has surfaced additional information concerning the variation in the UV performance of the human eye. He exhibits more UV sensitivity in the region between 350 and 400 nm than usually attributed by the data from a limited number of subjects by Said & Weald.

Figure 2 of the Norren & Vos paper suggests a variation, primarily in the slope of the absorption of the lens at wavelengths shorter than the break point near 420 nm based on “standardized” empirical data of their day. However, it does not address any variation in the ratio of aldehyde to alcohol in the lens. The data of Griswold & Stark is also primarily empirical data from clinical observations. Griswold & Stark due point to Stark (1987) and Boettner & Wolter (1962) for more fundamental data. The data of Boettner & Wolter is detailed and describes the difference in lens absorption among very young children and adults. It does not address the variation in lens transmission with respect to the aldehyde/alcohol ratio.

The bottom equation in **Figure 16.3.3-1** employed coefficients designed to match the curve of Griswold & Stark (solid line in **Figure 16.3.3-2**. However, the values of 0.1 and 1.0 are arbitrary and appear to vary significantly with the ratio of aldehyde to alcohol in the proteins of the lens. Robert M. (a trained physicist) has made observations with modern LED light sources and state of the art filters that suggest, his eyes reflect a very low percentage of the aldehyde relative to the alcohol in the proteins of his lenses, to the extent that his nominal short wavelength visual cutoff (dashed brown line intersecting with Rayleigh lines) due to his lenses is near 375 nm rather than the nominal value of 400 nm.

As in the case of most aphakics, Robert M. reports a desaturated chromatic perception near 400 nm (actually at 395 nm) that is described in this work as lilac. Following the construction described in **Section 17.3.3 .1**, the three desaturated chromatic perceptions of humans are:

---

[27]Boettner, E. & Wolter, J. (1962) Transmission of the ocular media. Invest. Ophthalmol. Vol. 1, pp 776-783

[28]Chen, D-M. & Stark, W. (1994) Electroretinographic analysis of ultraviolet sensitivity in juvenile and adult goldfish retinas. *Vision Res*. vol. 34, no. 22, pp 2941-2944

[29]Weale, R. (1954) Light absorption by the lens of the human eye, Optica Acta. vol . 1, pp. 107-110

- lilac at 395 nm (O = 0.00),
- azure at 494 nm (P = 0.00) and

- yellow at 572 nm (Q = 0.00).

As shown in **Figure 17.3.3-1**, best white for a normal human occurs at **W'**, for short wavelength irradiance with a centroid at 494 nm and long wavelength irradiance with a centroid at 572 nm. The contribution at W" is negligible under most illuminance conditions.

> The recent applied research into automobile headlights based on LED technology (but not phosphor enhanced technology) has clearly demonstrated that a perception of "white" cannot be obtained by using a single wavelength source of light. At least two narrowband sources must be combined to achieve an acceptable perception of "white" light by normal eyes.
>
> If only two LED wavelengths are selected, the reflected chromatic appearance of objects is greatly distorted by the lack of source energy in the broad spectral range required by most materials to provide an appropriate reflectance spectrum. As a result, automobile headlight research rapidly transitioned to phosphor enhanced LED sources.

For an aphakic human the situation is more complex. best white occurs at **W** corresponding to **W'** and **W"** both representing nulls in their corresponding difference channels.

Robert M. also describes a "violet-ish" chromatic perception at very short wavelengths. This work defines this chromatic perception at wavelengths shorter than 400 nm as "Monet," in honor of the first documented aphakic artist of the 19th Century, Claude Monet (1840-1926).

16.3.3.1.4 Macular absorption by the field lens

Over most of the surface of the retina in contact with the vitreous humor, the field lens exhibits negligible absorption. This is not true in the vicinity of the fovea. In this area, the relatively homogeneous nature of the field lens is disturbed by the movement of the cell nuclei associated with the signal manipulation stage out of the optical line of sight. This mechanism improves the scattering properties of the optical system. However, the bulk material of the INM appears to exhibit a higher absorption coefficient in the short wavelength region than the neural material it replaces. The result is a significantly higher optical absorption in the neural tissue overlaying the photoreceptors of the retina. Ruddock has collected data on this absorption[30]. The form of the absorption suggests a resonant condition in the molecules that make up the INM. In fact, the double peak in the absorption suggests a double resonance. This configuration is a text book case in electrical engineering[31]. An equation fitting this data very well is:

$$D = -\log\left(1 - \left[\frac{2.18}{\sqrt{1+\left(30\cdot\left(\frac{\lambda - 0.440}{0.440}\right)\right)^2}}\right]\cdot\left[\frac{1}{\sqrt{1+\left(30\cdot\left(\frac{\lambda - 0.487}{0.487}\right)\right)^2}}\right]\right)$$

where D is the measured optical density of the field lens at the fovea. The two expressions in the

---

[30]Ruddock, K. (1963) Evidence for macular pigmentation from colour matching data. Vision. Res. vol. 3, pp. 417-429

[31]Reference Data for Radio Engineers, 4th ed. (1957) NY: International Telephone & Telegraph Corporation. pg. 242-245

denominator are equivalent except for the fixed coefficients describing the two resonant wavelengths. The value of 2.18 can be separated into its square roots if it is desirable to make the numerators also equivalent. The numeric of 30 describes the resonance quality of each tuned circuit. 2Q=30. The only variable in the equation is the wavelength. This formula suggests that there are molecules present in the INM that exhibit resonances at 440nm and 487 nm. Early work by Wald[32] claimed "the human macular pigment was shown to be a carotenoid, apparently lutein or leaf xanthophyll." However, his data for xanthophyll showed three peaks at wavelengths longer than 400 nm. Ruddocks data does not show this feature. Whether the absorption is due to a single molecule is still open to question. Wald's original premise was the material was in the form of a surface coating of the retina. Later work has shown the situation to be more complex with the absorption taking place within the layers of the INM.

The graph of this function is presented in **Section 17.2.2**

16.3.3.2 The spectral sensitivity of the normal human

Figure 16.3.3-3 presents the best available data on the complete photopic luminous efficiency function of the human eye. It differs significantly from the C.I.E. Standard. It even goes beyond the suggested modifications by Judd by showing significant sensitivity down to 300 nm under suitable conditions. It also hints at Brezold-Brucke peaks at 487 and 580 nm. While Griswold & Stark described the curve as obtained under scotopic conditions using a 38 minute of arc diameter stimulus, this work suggests such a small stimulus necessarily measures the photopic characteristic of vision. See **Section 17.2.4.**

If the two peaks are due to the Brezold-Brucke Effect, the observed response was actually acquired under hypertopic conditions. Under these conditions, there is no shoulder in the long wavelength portion of the response to indicate the presence of the long wavelength spectral absorber. In addition, the peak absorption in the mid wavelength region is reduced due to saturation in the adaptation amplifier of that channel. As a result, the apparent sensitivity of the ultraviolet signaling channel is somewhat enhanced relative to the mid wavelength channel.

Griswold & Stark only made measurements down to 315 nm. Based on their data, the sensitivity of the eye may actually increase incrementally between 300 and than 315 nm. In either case, the overall response of the human eye shows a relative maximum in its response in the region of 300-340 nm.

Because of the very great dynamic range of the visual system, the sensitivity in the region between 300 and 330 nm may not be negligible under special conditions. As an example, the relative sensitivity of the eye in the ultraviolet may be significant in a discotheque featuring ultraviolet lighting.

Under intensity conditions similar to those employed by Griswold & Stark, the nominal coefficients for the normal human luminance equation are, C_{RU}' = 0000.3, C_{RS}' = 0.1, C_{RM}' = 1.000, and C_{RL}' = 0.1. As in the previous case, the prime denotes the coefficient is the product of the individual coefficients prior to the logarithmic term in the equation.

It is important to recognize the sensitivity of the ultraviolet spectral channel of the human eye is higher than that due to the S-channel in the region of 350 to 395 nms. This fact plays

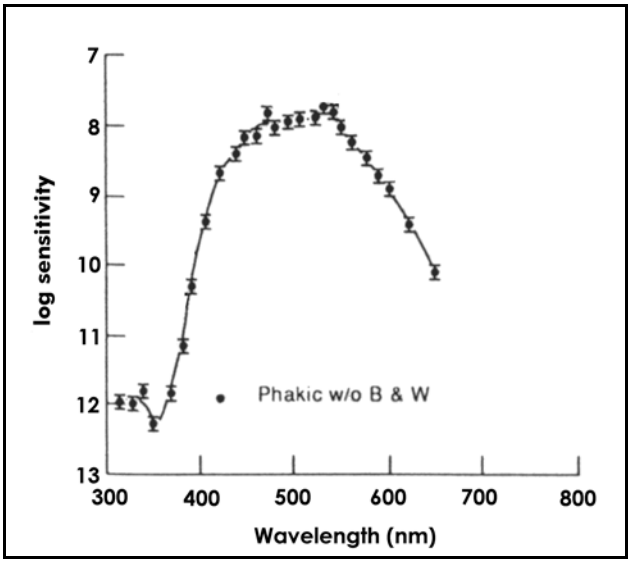


Figure 16.3.3-3 The complete luminous efficiency function for the human eye. There are hints of Brezold-Brucke peaks in this waveform. Data points from Griswold & Stark, 1992. The solid line is the function predicted by this model.

[32]Wald, G. (1945) Op. Cit.

a significant role in the chromatic performance of the human eye. While it may be acceptable to ignore the contribution of the U–channel to the perception of brightness for pedagogical purposes, it is not acceptable to discount completely the contribution of the U–channel to the perception of color. While the coefficient, C_{RU} may be set to zero in the Equations of Vision Matrix coefficient, the coefficient C_{OU} should not.

[xxx comment on CIE standard deviation per W&S.]

Professor Wright, one of the last living investigators who provided the underlying data, made an important comment concerning the CIE photopic spectral absorption function of 1924 in 1969. "When I look at the revised table of the x-bar, y-bar, z-bar, functions, I am rather surprised to say the least. You see, I know how inaccurate the actual measurements really were. (Laughter) Guild did not take any observations below 400 nm and neither did I, and neither did Gibson and Tyndall on the V-sub λ curve, and yet at a wavelength of 362 nm, for example we find a value y-bar of .000004929604! This, in spite of the fact that at 400 nm the value of y-bar may be in error by a factor of 10. (Laughter) "

16.3.4 The Chrominance Equations of vision

The Equations of Vision Matrix can be expanded to obtain the specific chrominance equations of vision occurring at the S-plane based on **Figures 16.2.1-1** and **16.2.1-2**. While the manipulations are similar, there is one unique difference.

> The difference between two logarithms is equal to the logarithm of the ratio of the arguments.

This critical difference, the difference in logarithms is concisely defined while the sum of logarithms is not, is critical to the operation of the chromatic subsystem of vision.

16.3.4.1 Experimental considerations

When developing the equations of chromatic vision, it is important to note two facts. First, the equations are developed to represent conditions at the S-plane of the retina. These equations are proposed to form a 1st order representation of what the biological system actually perceives. There are a myriad of reports in the literature of signals recorded within the LGN, the superior colliculus and area 17 of the occipital lobe. This discussion does not address the signals (some of great complexity) measured at these locations. Second, although less obvious in **Figures 16.2.1-1** & **16.2.1-2** than in the **Equations of Vision Matrix**, not all of the permutations of the signals at the pedicles of the photoreceptors are required to explain the perception of color. It has been found that a chrominance signal can be formed by the difference between any pair of spectral signals. Initially it was assumed that , except for a minus sign, the resulting pairs were commutative, A-B = B-A. However, it will be shown that this is not generally true following stage 3 of signal processing. Flicker experiments demonstrate a clear choice between these expressions within the visual system. See **Section 16.6.2.1**. Because of this preference, the signaling channels of vision used to transmit chrominance information are not symmetrical.

16.3.4.2 The general form of the chrominance equations

The general form of the three chrominance equations formed at the S-plane of the retina by horizontal cells can be defined as;

$$O(t) = \pm(m_{OU}\omega_U - m_{OS}\omega_S) \qquad \text{Eq. 16.3.4-1}$$

$$P(t) = \pm (m_{OS}\omega_S - m_{OM}\omega_M) \qquad \text{Eq. 16.3.4-2}$$

$$Q(t) = \pm (m_{OM}\omega_{M-} m_{OL}\omega_{L)} \qquad \text{Eq. 16.3.4-3}$$

It is not known which polarity of signal is used in the O–channel of human vision. However, it is known that the minus choice should be used in the P– and Q–channels, **Eq. 16.3.4-2** and **16.3.4-3,** when discussing flicker phenomena. When discussing a graphical presentation of color space, it is most useful to make a different selection of signs. For this purpose, the positive choice will be used in the O– and P–channels but

the negative choice will be used in the Q–channel. This last set of choices are those used in the matrix defined as the **Equations of Vision**. For consistency, this last choice will be used in the remainder of this section. It should be recognized that within the LGN and other engines of the visual system, many complex signal manipulations are undertaken. These manipulations may not correspond to those defined above.

Each of the above equations can be expanded to show their logarithmic relationship to an external stimulus.

$$O(t) = m_{OU}C_U \text{ Ln } G_U\, \sigma_U(t) - m_{OS}\, C_S \text{ Ln } G_S\, \sigma_S(t) \qquad \text{Eq. 16.3.4-4}$$

$$P(t) = m_{PS}C_S \text{ Ln } G_S\, \sigma_S(t) - m_{PM}\, C_M \text{ Ln } G_M\, \sigma_M(t) \qquad \text{Eq. 16.3.4-5}$$

$$Q(t) = m_{QL}C_L \text{ Ln } G_L\, \sigma_L(t) - m_{QM}\, C_M \text{ Ln } G_M\, \sigma_M(t) \qquad \text{Eq. 16.3.4-6}$$

16.3.4.3 The chrominance equations in the photopic region

Using the logarithmic identity noted earlier, and restricting the discussion to the photopic region, the product of the coefficients in front of each logarithmic term are essentially equal. Each equation can be factored to remove these coefficients. The resulting equations become;

$$O(t) = \text{Ln } [G_U\, \sigma_U(t)/\, G_S\, \sigma_S(t)] \qquad \text{Eq. 16.3.4-7}$$

$$P(t) = \text{Ln } [G_S\, \sigma_S(t)/\, G_M\, \sigma_M(t)] \qquad \text{Eq. 16.3.4-8}$$

$$Q(t) = \text{Ln } [G_L\, \sigma_L(t)/\, G_M\, \sigma_M(t)] \qquad \text{Eq. 16.3.4-9}$$

These are the chrominance channel potentials formed at the output of the horizontal cells in the S-plane prior to passing to the ganglion cells. Where the numerators and denominators in one of the equations are equal (indicating a balance in the signals presented to the horizontal cells), the signal level for that channel is equal to zero.

16.3.4.4 Color constancy within the photopic region

As discussed earlier, the purpose of the adaptation process is to maintain a constant average signal amplitude at the output of each spectral channel. This is accomplished by maintaining the product of the average signal level, σ_k times the gain, G_k, equal to a constant. As in **Section 16.3.2.1**, let the signal in each spectral channel be defined as the absolute modulation around an average value, σ_k. The value of this varying signal is then σ_k times $\text{mod}_k(t)$. Substituting these expressions in the above equations,

$$O(t) = \text{Ln } [G_U\, \sigma_U\, \text{mod}_U(t)/\, G_S\, \sigma_S\, \text{mod}_S(t)] = \text{Ln}[\text{mod}_U(t)/\text{mod}_S(t)] \qquad \text{Eq. 16.3.4-10}$$

and following the same procedure,

$$P(t) = \text{Ln}[\text{mod}_S(t)/\text{mod}_M(t)] \qquad \text{Eq. 16.3.4-11}$$

$$Q(t) = \text{Ln}[\text{mod}_L(t)/\text{mod}_M(t)] \qquad \text{Eq. 16.3.4-12}$$

These equations say that the signal in the chrominance channels under photopic conditions are independent of the level of any slowly changing input irradiance regardless of its spectral content. **This is the methodology leading to Color Constancy**.

Note the phase of the signal in each chrominance channel will be a function of which spectral channel is being modulated.

16.3.4.5 Detailed equations of the chrominance channels

This section concentrates on the P– & Q– chrominance channels. The equations of the O–channel are similar to those of the P–channel..

Careful examination of the above equations show they are not completely independent, particularly when the system is excited by narrow band stimuli. The consequences of this will be examined in **Section 16.3.4.5.1**.

The equations of the previous sections did not recognize the fact that the individual neurological circuits are connected directly (without a capacitor or transformer separating them). The resulting circuits are called direct-connected, or DC, circuits. To completely understand such circuits, it is necessary to recognize the fact that a bias applied to one circuit may generate a potential at the output of that circuit without any input stimulus. To completely understand the operation of such circuits, it is necessary to define both the exact direct current and alternating current representation of their output potentials.

Errors in the bias potentials of the various neurons of the retina lead to significant abnormalities in color vision of the type documented extensively by Lakowski. These errors are reviewed in **Section 18.1.5**. The equations involved will be presented in **Section 16.3.4.5.2**.

The mere presence of a bias in the case of the lateral cells is important to the overall operation of the chrominance, and other lateral, channels. It accounts for the gradual loss of color perception as the stimulus intensity is reduced and the fact this loss is chromatically asymmetrical. The equations explaining this situation will be presented in **Section 16.3.4.5.3**.

16.3.4.5.1 Cross coupling between P– and Q–channel signals

Expanding the equations for the P– and Q–chrominance channels within the photopic operating region (all adaptation amplifiers are within their operating range associated with variable gain), the following equations result.

$$P(t) = m_{PS}C_S \text{ Ln } G_S \text{ Integral } E(\lambda)\, S^x(\lambda)O(\lambda)\, R_S(\lambda)\, d\lambda - m_{PM}\, C_M \text{ Ln } G_M \text{ Integral } E(\lambda)\, S^x(\lambda)O(\lambda)\, R_M(\lambda)\, d\lambda \quad \text{Eq. 16.3.4-13}$$

$$Q(t) = m_{QL}C_L \text{ Ln } G_L \text{ Integral } E(\lambda)\, S^x(\lambda)O(\lambda)\, R_L(\lambda)\, d\lambda - m_{QM}\, C_M \text{ Ln } G_M \text{ Integral } \int E(\lambda)\, S^x(\lambda)O(\lambda)\, R_M(\lambda)\, d\lambda \quad \text{Eq. 16.3.4-14}$$

With the adaptation amplifiers operating, the average value of the terms $G \text{ Integral } E(\lambda)\, S^x(\lambda)O(\lambda)\, d\lambda$ will approximate 1.00. If one attempts to measure the spectral response of the P– and Q–channels using a field of uniform reflectance and a monochromator as a source, Sp(λ), of constant peak intensity, the equations can be written as;

$$P(t) = m_{PS}C_S \text{ Ln } Sp(\lambda)R_S(\lambda) - m_{PM}\, C_M \text{ Ln } Sp(\lambda)R_M(\lambda) \quad \text{Eq. 16.3.4-15}$$

$$Q(t) = m_{QL}C_L \text{ Ln } Sp(\lambda)R_L(\lambda) - m_{QM}\, C_M \text{ Ln } Sp(\lambda)R_M(\lambda) \quad \text{Eq. 16.3.4-16}$$

As the monochromator wavelength becomes very short, these equations reduce to;

$$P(t) = m_{PS}C_S \text{ Ln } Sp(\lambda)R_S(\lambda) \;\&\; Q(t) = 0.00 \quad \text{Eq. 16.3.4-17}$$

This condition forms one asymptote of the Young-Helmholtz axes in the New Chromaticity Diagram for Research. It overlays the short wavelength terminus of the one Hering axes.

As the monochromator wavelength becomes very long, these equations reduce to;

$$P(t) = 0.00 \;\&\; Q(t) = m_{QL}C_L \text{ Ln } Sp(\lambda)R_L(\lambda) \quad \text{Eq. 16.3.4-18}$$

This condition forms another asymptote of the Young-Helmholtz axes in that diagram. It overlays the long wavelength terminus of the other Hering axes. Under the third condition, where the absorption by the M–channel absorber is dominant, these equations reduce to;

$$P(t) = Q(t) = -\, m_{PM}\, C_M \text{ Ln } Sp(\lambda)R_M(\lambda) \quad \text{Eq. 16.3.4-19}$$

This condition represents a third asymptote of the Young-Helmholtz axes in the New Chromaticity Diagram

for Research. It is not convergent with any of the termini of the Hering axes. The values of P & Q are not independent in the green region. At 532 nm, they are equal.

The existence of these three distinct asymptotes in the color space of the New Chromaticity Diagram for Research illustrate the conflict between color presentations related to Young-Helmholtz and those related to Hering. When defined as in this work, the Hering axes are orthogonal. The Young-Helmholtz axes are not. The situation becomes more complex if the O–channel is also considered.

16.3.4.5.2 Color abnormalities due to neuron bias errors

To describe the complete signal within a series of DC coupled circuits, it is necessary to account for both the dynamic signal amplitudes and the pedestal heights associated with those signals. To do this, the notation of [**Figure 16.2.1-2**] will be expanded to include a pedestal amplitude in the absence of any signal. The quiescent potential at the output of each neuron will be described by a nominal constant, H_{XQ}. Thus, the quiescent potential at the pedicle of the S-channel photoreceptor would be H_{SQ}.

In progressing through the circuitry of the lateral cells, it is seen that the synapses leading to each input impedance of both neurites of the lateral cell Activa contains a diode. Because of the transfer characteristic of these diodes may be different, the relative amplitudes of the spectral signal components applied to the Activa within each lateral cell may not be the same. Furthermore, since these circuits are DC coupled, the output signal from the lateral cells may exhibit a DC component , H_{XQ}, that will also be presented to the circuits of the following stage(s). The exact value of the DC component at different points in the retina will require laboratory measurements. As shown in Figure xxx, the nominal value for the output of a lateral cell is -75 mV relative to the matrix surrounding the cell. There is also a gain coefficient associated with the podite input structure of the lateral cell but it can be assumed to be one. Based on [**Equations 17.3.3-5** & **17.3.3-6 xxx**] , the complete form of P and Q, under both photopic and mesotopic conditions, can be written as

$$P(t) = H_{PQ} + m_{PS}C_S \, Ln\,[H_{SQ} + G_S\,\sigma_S(t)] - m_{PM}\,C_M\,Ln[H_{MQ} + G_M\,\sigma_M(t)] \quad \text{Eq. 16.3.4-20}$$

$$Q(t) = H_{QQ} + m_{QL}C_L\,Ln[H_{LQ} + G_L\,\sigma_L(t)] - m_{QM}\,C_M\,Ln[H_{MQ} + G_M\,\sigma_M(t)] \quad \text{Eq. 16.3.4-21}$$

These equations are important because they determine the potential applied to the ganglion cells of Stage 3. The ganglion cells can accept only a limited voltage range at their inputs while staying within their operating range. Any abnormality in the quiescent potentials in the above equations can cause abnormalities in the perceived color performance of the visual system. As mentioned above, these quiescent signal errors are the primary source of the abnormalities documented by Lakowski.

16.3.4.5.3 Loss of color perception at reduced stimulus intensity

As in the above paragraph, introducing the quiescent DC potential at the output of the horizontal cells illustrates another phenomenon of vision, the loss of chromatic sensitivity at low light levels.

Figure 16.3.4-1(a) & **(b)** present plots of the signal component of P and Q as a function of the wavelength for a monochromatic stimulus. The amplitudes of the S-, M- and L-channel signals are shown parametrically. Under photopic and hypertopic conditions, the adaptation amplifiers maintain a constant maximum output amplitude in each channel. They also maintain the same amplitude relationship between these three photodetection channels on short term basis. This relationship insures that the P and Q signals are not affected by changes in the illumination level. The heavy lines in **(a)** and **(b)** represent both the hypertopic and photopic signal levels in the P and Q channels.

Within the mesotopic illumination range, the signal level applied to the adaptation amplifiers becomes so low that the amplifiers are no longer able to maintain the constant maximum output amplitude at the pedicle of each photoreceptor. Under this condition, the amplitudes of both P and Q begin to decrease toward the null value . The dotted lines represent the response in the mesotopic region a factor of two below the transition from the photopic region. The dashed lines represent the response in the mesotopic region a factor of five below that transition. The decrease in response is symmetrical in the P channel. In the Q channel, the decrease is less symmetrical because the square law function due to the translation process is no longer masked by the variable gain characteristic of the adaptation amplifier. In the Q channel, there is an indication that the predicted chrominance response falls faster in the shorter wavelength region than it does in the long wavelength region.

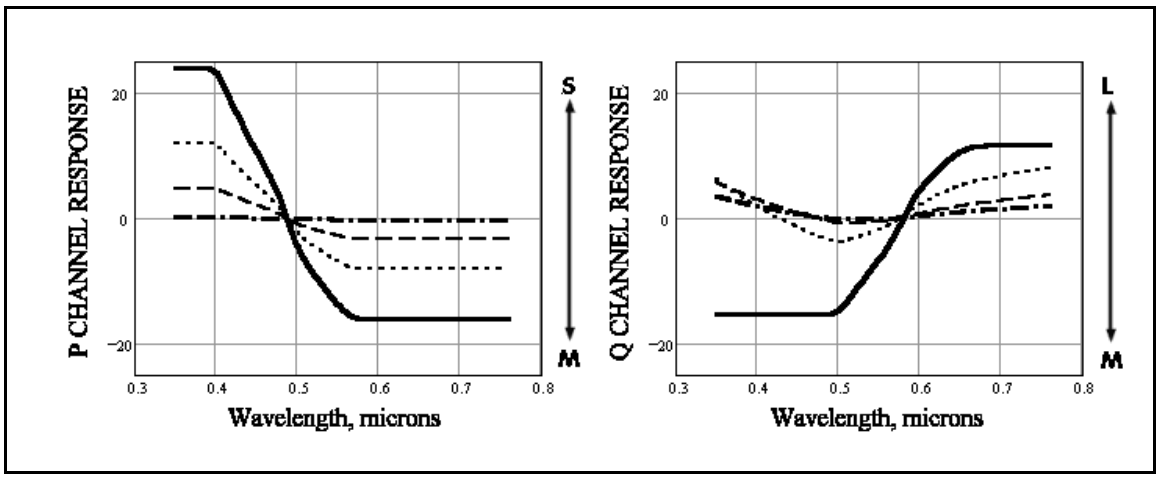


Figure 16.3.4-1 Response of the chrominance channels versus an equal integrated flux monochromatic stimulus following recent adaptation to an equal flux per unit wavelength background. Solid line represents the hypertopic and photopic regions in both graphs. Dotted line represents the mesotopic region at a factor of 2 below the bottom of the photopic region. The dashed line is a factor of 5 below the bottom of the photopic region. The dash-dot line represents the scotopic condition. Note the more rapid, and asymmetrical decrease in amplitude of the Q channel signal compared to the P channel signal. Left hand scales are linear but the values are irrelevant.

As the signal levels received from the photodetection channels approach zero in amplitude, the P and Q channels no longer provide any chrominance information to the remaining portion of the signaling system. The dash-dot lines in the figure illustrate this condition. Under these conditions, the total signal delivered to the ganglion cells consists of a quiescent value for P and Q. These quiescent potentials are not zero. It is these potentials that cause the quescent stream of action potentials at the output of each midget ganglion cell associated with a differencing type signaling channel. These pulses have a typical spacing of 33 ms (30 Hertz).

Afeature of the signal projection technique used in Stage 3 is that the nominal interval between the action potentials of each data stream are discarded in recovering the signal information. This insures that the signals recovered by the stellate cells of the brain will not indicate a chromatic shading of a scene due to an error in the quiescent potential of one of the preceding signaling circuits. However, the dynamic range of the recovered signal may be limited (generally asymmetrically) by the limited input dynamic range of the ganglion cells discussed above. This limiting is also apparent in the data of Lakowski.

16.3.4.5.4 Creating a color continuum from the P– & Q–channels

In developing a New Chromaticity Diagram for Research in **Section 17.3**, (see **Figure 17.3.3–6**) a problem arises in connection with the above figure (and the equivalent pair of frames for the O & P channels). There is no logical place in color space for P and Q values more negative than –7 (in the arbitrary units used above), yet the graphs indicate values as low as –15 when the P and Q values are computed individually. It would appear that such large negative values would lead to confusion or ambiguity in the perception of color in the brain. Several approaches to avoiding this problem have been explored. The most attractive approach from the neurological perspective appears to be to limit the negative excursion of both the P and Q signals under all conditions. This limiting could be performed anywhere in the visual system after P and Q signal formation and before the information was used to create a saliency map input signal. A simple way to achieve this performance is to have the stellate decoder output circuit (or the following input circuit) be limited in its negative excursion (based on the sign convention and units used here). The stellate decoder has a natural asymptote controlled by the highest frequency of the action potential stream and the time constant of the output circuit of the decoder. The voltage of this asymptote can be modified further by an electrical bias circuit. The limiting could also be performed by the following input circuit. In the case of the following input circuit for either the P or Q channel, it would be biased to cutoff near –7 units. This approach would provide a sharp limiting function near –7 units. Alternately, the circuit could be reconfigured so the following output circuit went into saturation for high negative values. This would provide a softer limiter function. Any of these approaches would work under all spectral conditions.

A similar description of the circuitry associated with the O and P channel pair can be developed. However, under some conditions, it would be necessary to replace each P–channel in the optic nerve with both a P–channel and a reciprocal P–channel in order to create a complete conformal and unambiguous physiological chromaticity space. Further study of the database is needed to determine the precise signaling plan and decoding approach used in tetrachromatic vision.

Data is accumulating that suggests the above limiting is associated with the Stage 3 signal projection circuits. It is now known that the M–channel signal drives the midget ganglion encoders to maximum frequency in both the P and Q channels. When coupled with the time constant of the stellate decoder circuits, these two parameters form a natural limit to the output signal amplitude of the overall circuit. Whether this asymptote is approached at the proper rate under this proposed operating mode is yet to be determined. Alternately, the maximum frequency of the action potential stream generated by the midget ganglion neurons may form the limiting asymptote in the overall operation. See **Section 7.2.2**.

16.3.4.5.5 An anomaly confirming the tetrachromatic concept

While studying tetartanopes in the nomenclature of this work (tritanopes in his nomenclature), Wright presented a figure showing poor operation of the P–channel in color discrimination[33]. Of equal interest was the fact he showed excellent discrimination in the O–channel region for a number of subjects. **Figure 16.3.4-2** presents a portion of his data. It can be compared to the differentiated form of **[Figure 17.3.3-2**]. Note the loss of color discrimination within the region of 437-532 nm in both the left and right frames. This poor performance in the region normally dominated by the P–channel is indicative of a true tetartanope. There are two other interesting points in this figure. Note the significantly poorer performance of observer B on the right in the area of 480–575 nm. This is suggestive of a bias error of the type explored by Lakowski.

Of greater significance is the surprisingly good color discrimination in the area of 400 to 450 nm. This is the region of color perception expected to be dominated by the O–channel of the chrominance system. While this channel is not normally associated with human vision, its presence was projected based on the work of Stark & Griswold and of Tan discussed in **Section 17.2.2**. This performance documented 50 years ago by Wright provides additional evidence that the human visual system is tetrachromatic, with its ultraviolet performance blocked by the size of its own lens system.

---

[33]Wright, W. (1952) The characteristics of tritanopia *J Opt Soc Am* vol. 42, pp 509-521

16.3.5 The appearance equations of vision

It is difficult to ascertain direct relationships relating to the spatial integration properties of the human visual system because of the nature of the experiments usually reported. There are a great number of experiments, including those of the center-surround difference type. However, most of them have used one of two main environments; a white or other poorly defined light source, or two different colors that have not been chosen to excite specific chromophores. The results of these experiments are characterized by the presence of uncontrolled parameters and less than optimum signal to background conditions. Repeating many of these experiments after a redesign of the procedures used could lead to much more useful information. In the meantime, only general conclusions can be drawn.

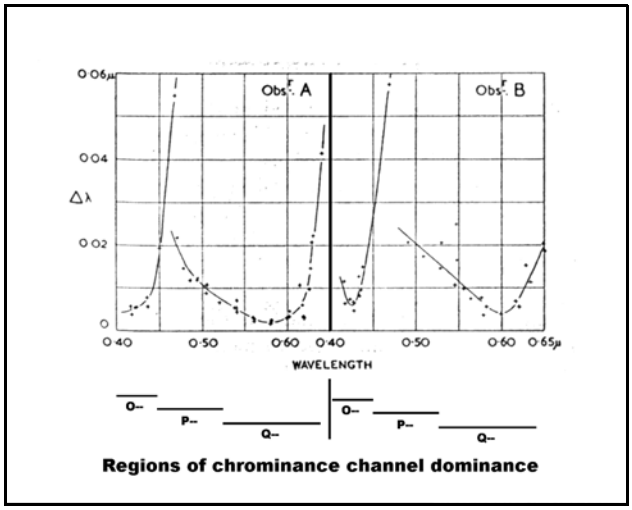


Figure 16.3.4-2 Color discrimination versus wavelength for two tetartanopes (tritanopes under earlier terminology). Note the loss of P–channel performance in the region between 437 and 532 nm. Note also the loss in performance near 480-550 nm in Observer B. This could be indicative of some degree of weakness in the M–spectral channel or its coupling to the Q–channel. Data from Wright, 1952

The appearance signaling channels, the G–channels, take on two distinct characters in vision. There are those signal related to the correlation of scene content outside of the foveola. These are primarily processed in the lateral geniculate bodies. Besides analysis of scene content, the lateral geniculate bodies are also responsible for scene convergence in a binocular visual system. There are separate and distinct circuits, associated with the G'–channels, responsible for correlation of scene content within the foveola and the detailed oculomotor operation involved in analyzing and tracking of that content.

There have been three classes of experiments designed to evaluate the spatial performance of vision. There are the center surround type of experiments using circular test stimuli. There are experiments using a one dimensional grating as a stimuli and there are experiments using two-dimensional gratings as a stimuli. Recently, experiments have attempted to address the theoretical mathematical problems associated with using a finite grating that can be described mathematically as a grating of infinite extent windowed by a clear aperture of a finite size. The recent tests have introduced what are defined as Gabor patches after Professor Gabor (1900-1979) of University College, London. These patches use a window of variable opacity, usually Gaussian in character along either one or both axes. While these patches tend to reduce spatial transients associated with the signaling, they also significantly reduce the ability of the correlation process used within the visual system related to the foveola. A more sophisticated form of Gabor patch employs a Checkerboard pattern windowed by a Gaussian or other pattern as the stimulus.

While center surround experiments are easy to implement in the laboratory, they offer limited insight into the operation of the visual system. Even with more care in experiment design, their potential is low. The use of one dimensional gratings offers more potential insight. By rotating the grating, it is possible to determine axes processed preferentially in a given species of animal. It is also possible to sweep the windowed grating across the field of view to ascertain any velocity preferences associated with the animal. The Hubel team has collected good data in this area.

When using patterns more complicated than simple gratings, care must be taken to differentiate between the signal analysis performed in the "hard-wired" part of the visual system and the "soft-wired" part. The former is usually thought of as involving perception of previously recognized (and thereby learned) shapes whereas the latter is more involved with the cognition process used to identify new, and particularly

complex, shapes. There is reason to believe most of the learned patterns are stored in the Cerebellum while the learning of new patterns requires the participation of the occipital cortex.

16.3.5.1 Spatial signal processing related to the extra-foveal retina

The human visual system does not appear to make intense use of spatial signal processing for object space imaged outside of the foveola. The human retina is very sparsely populated by amercine and other lateral cells generally associated with the second lateral matrix of the signal processing system (See **Figure 15.2.3-1**). For the capabilities of other chordates, the reader is referred to the work of the Hubel team, and the xxx teams.

16.3.5.1.1 The mechanism of spatial processing in the retina

There has long been discussion in the literature as to the mechanism of spatial processing within the visual system. Most of the discussion has been conceptual. Detailed models of the actual process have been rare. The primary purpose of spatial signal processing within the retina (a portion of the "hard-wired" part of signal processing) is to provide alarm signals. The second purpose is to provide signals to the oculomotor subsystem designed to bring the image of the threat onto the highest acuity portion of the field of view, the foveola.

The overall architecture of spatial signal processing within the retina does not involve processing in the spatial domain. The primary processing relies upon time delays between the signals arriving at either summing or differencing circuits from individual photoreceptors. It is the arrival of signals with different time delays that are used to generate alarm signals. The variation in time delays can be used to determine motion and direction of motion within the field of view. Only an edge is required within the field of view in order to extract these parameters. However, if the scene contains a grating of specific pitch, the accumulated signals can exhibit a considerably higher signal to noise ratio. Such signals can induce an alarm signal at considerably lower signal intensity levels relative to a fixed threshold.

The time delay associated with the signal from one photoreceptor to one spatial signal processing neuron is dominated by the velocity of analog electrical signals within the dendrites of the neuron. These velocities are quite low compared to the velocity of action potentials. Alternately, the length of the dendrites can be significant. The time delay associated with a given dendrite can be calculated by dividing the dendrite length by the signal velocity. At a larger scale, the signals from different areas of the retina exhibit significantly different time delays upon reaching the LGN. These delays are illustrated in **Figure 15.2.4.3** where the paths of the different neural outputs are described by Reyem's Loops. These loops are the compliment of Meyer's loops that are found between the LGN and the occipital cortex. They also perform a complimentary function. Reyem's loops disperse the data in time in order to aid the LGN in determining position and relative motion within object space. Meyer's loops tend to remove the dispersal introduced by Reyem's loops prior to further processing in the occipital cortex.

Both the LGN and the cortex rely upon two dimensional cross correlation of signals between spatially related signaling channels. The mathematics of the correlation processes carried out within the individual engines of these bodies are not currently understood.

16.3.5.2 Spatial signal processing related to the foveola

It is the imagery formed on the foveola that is of primary importance in animals. The higher the animal along the evolutionary chain, the more important the analysis of detail in the scene becomes. This is particularly noticeable with regard to eating habits. Most animals will consume whole anything they can capture. However, many animals learn to discard certain organs associated with their prey. To aid in this process, some animals learn to wash their food in order to discern more carefully the organs they prefer or abhor. It is the appearance channels of the visual system that support this type of object analysis.

16.3.5.2.1 The human appearance channel associated with the foveola

This material supports **Section 17.7.5**.

As highlighted in **Figure 16.2.1-1**, and in more detail in **Figure 15.2.4-3**, the signals from the photoreceptors within the foveola are passed directly to the pretectum of the midbrain. This body is a major part of the Precision Optical System (POS). The POS contains a large number of individual signal processing engines. The algorithms used by these engines are not currently known. In some cases, the

scope of these algorithms can be described.

The signals from the foveola are passed along the G'–signaling channel via a specific bundle of axons within the optic nerve. The axons of this bundle terminate in the pretectum. The signals from each foveola are presented to the pretectum separately. The pretectum initially merges these signals based on parallax data passed to it from the LGN. For the purpose of image analysis, it appears that each signal is treated without regard to its spectral lineage. It is possible that two signals from different eyes may be summed even though they originate from different spectral channels. This is a virtual necessity because of the parallax problem and the capabilities of the oculomotor system.

Following merging, the signals are presented to a two dimensional signal correlator. This two-dimensional correlator of the pretectum is responsible for the analysis of fine detail in an image. It ignores the spectral origin of the signals presented to it. The correlator operates in a binary mode with relation to signal amplitudes. It is primarily interested in the intensity of the individual signals at each spatial position within the foveola. The intensity is either above or below a threshold value.

This correlator is not based on serial signal processing. It can be thought of as designed to determine the correlation coefficients along each row and column of the matrix of signals presented to it. This matrix is contains approximately 23,000 elements in a circular formation with diagonals of about 175 elements. This array represents a circular area of about 0.35mm on the retina or 1.18 degrees in object space.

The specific characteristics of this two-dimensional correlator can only be given conceptually. The output is a set of coefficients describing the spatial content of the image projected on the foveola without regard to chromatic content. While there is probably a second set of algorithms that further compacts this information, little can be said authoritatively. It is clear that the image presented to the foveola is usually scanned in very small increments using the POS to generate microsaccades associated with tremor. This scanning process generates several "views" of the same pattern. By correlating the data set from these views, a higher signal to threshold ratio is available for later analysis. The final set of coefficients are used within the brain to analyze scene content. This set of coefficients are transmitted to Area 7 of the cerebral cortex (not area 17) via the pulvinar pathway. They are placed in the visual portion of the saliency map of the brain (**Section 15.2** and **Figure 15.2.4.3**). This map is primarily a short term memory for the collection and analysis of new sensations. As part of this map, they are analyzed by the engines of the cognitive centers located in the frontal lobe and disposed of appropriately. They are also transmitted to the Cerebellum for permanent storage on instructions from the cognitive centers. When stored, they constitute the permanent memory of what the animal has analyzed. They form a major part of its ability to recognize scenes in the future. In humans, the cerebellum is the key to not only the recognition of natural patterns but the recognition of symbolic patterns. The recognition of these symbolic patterns is the key to reading. This capability is discussed in **Section 17.7.5**.

The equations of oculomotor performance associated with tremor are discussed in **Section 7.3** through **7.5**.

16.4 The temporal equations of vision

The complete temporal equations of vision are quite complex and includes many variables in the general case. The general case includes all of the parameters and constraints of both the signal detection, signal manipulation, and signal projection subsystems. This complexity reaches the level of awkwardness unless the equations are subdivided. A convenient special case applies to warm blooded animals only. This simplification removes a complex temperature related term from the equation. A second simplification is to ignore the transient aspects of the equation related to the quantum mechanical process of photodetection. This simplification removes a complex illumination related term from the equation and leaves only those terms with a time constant longer than 30-50 milliseconds. However, it still includes the asymmetrical characteristics of the signal projection system. These are the characteristics that account for many of the special chromatic effects associated with moving or transient scenes. If these asymmetries are also factored out, the residual equation is the adaptation characteristic of vision, both dark and light, as it is known in the literature.

Tranchina has recently provided a survey of the state of transduction in accordance with the common

wisdom based on biochemical mechanisms[34].

> It was refreshing to see the word Calculus used in the title of his paper, even if used in a generic sense. Similarly, he pointed to a related paper that actually employed a "rigorous approach" employing theorems, lemmas and proofs. Unfortunately, the data supporting that paper was acquired under *in-vitro* and apparently non-functional conditions since saturation occurred at such low stimulus levels[35].

His last paragraph enumerates some of the open questions based on that approach at that time. He did not address the applicability of the analyses if there is no membrane surrounding the outer segment. Most of these open questions do not appear in the alternate and much simpler electrolytic mechanisms described in this work.

This section will be subdivided in order to discuss these simplified cases. A distinction will be made between the temporal aspects of signal generation, signal manipulation and signal transmission. For this section, the operation of the iris and the operation of the adaptation amplifier will be treated in the section on signal manipulation.

16.4.1 Temporal equation of signal generation in stage 1

16.4.1.1 The Photoexcitation/De-excitation, P/D, Equation

The short term temporal characteristics of signal generation are entirely within the domain of the P/D equation associated with each individual photoreceptor cell. The signal generation function of the P/D equation normally operates at an efficiency as a function of illumination that is very close to 1.0. When developed using complex algebra and an impulse input function, the P/D equation is seen to consist of two parts, a delay term and a transient amplitude term. Both of these terms are functions of the input intensity. The complete derivation of this equation is presented in **Appendix A**.

Hodgkin, et. al. began trying to explain the photoexcitation/de-excitation process during the 1960's. They began with a fundamental first-order differential equation of the type shown in Eq. 16.4.1-1[36].

$$\tau \bullet (dq/dt) + q = Q_2 \bullet q_f \qquad \text{Eq. 16.4.1-1}$$

where q is the number of excited electrons at any time, q_f is the time dependent forcing function driving the solution, Q_2 is a constant term related to q_f and τ is the time constant of the process. It is the difference in the forcing function, q_f, that provides the unique solution to the P/D Equation. With this forcing function, the delay in the resulting response is seen to vary continuously as a function of excitation. Hodgkin, and other earlier investigators have used a simpler forcing function that did not reflect the details of the differential processes involved. The investigators resorted to multi-stage fixed (and linear) filters of the resistor-capacitor (RC) type to emulate the empirically measured performance. The number of stages in these filters varied from 1 to 30 with each investigator picking a number that best matched his data over a limited range. The results were unsatisfying (see fig 3 of Fourtes & Hodgkin). The frustration of Fourtes & Hodgkin with this multi-stage approach can be seen in their comment; "The introduction of isolating units is admittedly highly speculative, but it has the merit of generating simple equations. . ." The same approach was used in a subsequent series of papers by Baylor, Hodgkin & Lamb[37]. Many others have attempted to

---

[34]Tranchina, D. (1998) The Calculus of rod phototransduction. *J. Gen Physiol*. vol. 111, pp 3-6

[35]Nikonov, S. Engheta, N. & Pugh, E. (1998) Kinetics of recovery of the dark-adapted salamander rod photoresponse. *J. Gen Physiol*. vol. 111, pp 7-37

[36]Fourtes, M. & Hodgkin, A. (1964) Changes in time scale and sensitivity in the ommatidia of *Limulus*, *J. Physiol*. vol. 172, pp 239-263

[37]Baylor, D. Hodgkin, A. & Lamb, T. (1974) The electrical response of turtle cones to flashes and steps of light. J. Physiol. Vol. 242, pp 685-791 (three individual papers)

empirically fit the multistage RC filter approach to their data[38].

Later, Lamb went on to seek a partial solution to the P/D problem by only addressing the leading edge of the *a*-wave as measured via the ERG. Initially using a Michaelis Equation, he transitioned to what he describes as a "delayed Gaussian" function. Like the Michaelis Equation, this integral of the Gaussian Distribution is monotonic[39]. It is unable to account for the fall in the response of a photoreceptor following the initial rise due to an impulse stimulus.

Cideciyan & Jacobson have noted the poor fit between their data at high stimulus levels and the proposed delayed Gaussian model of Lamb, et. al. They give that model the label model 1. They note the model (1) describes families of waveforms with a "constant latency and an increasing slope." In contrast, they define their model (2) "that attains a "constant slope and a shortening latency" with stimulus amplitude. The difference hinges on the fact that they are both using piece-wise empirical formulas to fit a single continuous (but quite complex) function. Both the Lamb and Cideciyan approaches are based on the kinetics of multiple processes that are not all understood. They have both settled on a three-stage RC filter as a fundamental configuration within their overall model. No functional model related to these three filter stages has been provided. It will become apparent, using a more detailed model, that Cideciyan & Jacobson are comparing apples and oranges. The Lamb group are measuring currents from a single photoreceptor with a suction pipette (a Class A waveform of this work, see **Section 11.1.5**) while Cideciyan & Jacobson are measuring an integrated voltage from a large number of cells at a point remote from the sources. The complete waveform they measured is quite complex. The initial part of their ERG waveform is defined as a Class C waveform in this work (see **Section 16.7**). The Class C waveform is more precisely defined than, and is a component of, the *a*-wave of the ERG. While the Lamb group is measuring and modeling a first order differential process, Cideciyan & Jacobson are measuring and modeling a second order differential process due to their instrumentation. The limitation on the slope found by Cideciyan & Jacobson probably relates to the bandpass of their instrumentation.

Both of the above models assume a putative photoisomerization of rhodopsin and some sort of gate arrangement controlling the current flow through a putative membrane surrounding the outer segment. Neither of the above assumptions are supported by this work. Whereas the early work of Fourtes & Hodgkin involved electrophysiological probes, the later investigators employed the electroretinograph. The resulting recordings are significantly different as will be shown in **Section 16.7 xxx**].

Figure 16.4.1-1 illustrates the fundamental situation as interpreted in this work. The equation presented here relies upon photoexcitation of electrons into an excited state (without relying upon any mechanical rearrangement of the participating molecules) and their subsequent de-excitation closely coupled with the generation of free electrons within the dendritic structure of the neural portion of the photoreceptor cells. The unexcited molecules or one of the Rhodonines are shown at the bottom. The first step is shown on the left of the figure. As a result of the excitation of an electron from the n– or ground state, it appears in the σ energy band of the same molecule. Since the molecule is part of a liquid crystalline structure in contact with the dendrites of the photoreceptor cell, the excited electron is able to physically travel through the crystal until it reaches the location of the appropriate neural surface. At this surface, the excited electron participates in a "handshake" with an unexcited electron within the neural tissue. As a result of this process, the excited electron is returned to its ground state and the associated molecule is returned to its unexcited state. This process is shown on the right. Rather than show a more complicated drawing, the handshake is shown symbolically as creating an exciton that they impacts with the neural tissue. However, the actual process requires the simultaneous deexcitation of the Rhodonine molecule and the creation of a free electron in the neural tissue. There is no need for the physical movement of any of the light sensitive molecules within the outer segment or within the IPM. However, there is movement of both excited and ground state electrons within the liquid crystal. This movement involves a finite time. This time must be represented in the overall process because it impacts the number of ground state electrons available for absorption by the liquid crystal. All of the above kinetic activity is at the quantum electronics level rather than the molecular chemical level as assumed by the above authors.

---

[38]Van de Grind, et. al. (1973) Temporal transfer in the visual system. *In* Handbook of Sensory Physiology, Vol VII, No. 3, Jung, R. *ed*. NY: Springer-Verlag, pp 532-538

[39]Pugh, E. & Lamb, T. (1993) Amplification and kinetics of the activation steps in phototransduction. *Biochem. Biophys. Acta* vol.1141, pp 111-149

To properly treat the time delay related to the travel of the excited electrons requires the use of complex algebra to solve the appropriate differential equation. The solution to this equation is a continuous complex expression defining the complete photoexcitation/de-excitation process. The expression can be separated into an amplitude function and an intimately associated phase (or time delay) function. For simplicity, this separation will be employed below. The complete development of the P/D Equation discussed below will be found in Appendix A.

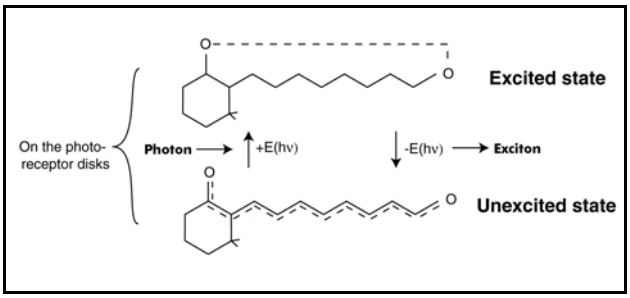


Figure 16.4.1-1 The fundamental mechanisms of photoexcitation/de-excitation.

The development of the Photoexcitation/De-excitation Equation based on the above mechanisms uses the same second order differential equation as shown above. However, the initial conditions and the forcing function are different. In particular, the forcing function involves the number of ground state electrons available at any instant for excitation. This number depends on the lifetime of the excited electrons that are removed from the available pool. That lifetime depends on the physical dimensions of the outer segments and the velocity of the excited electrons traveling over the surface of the individual disks. This introduction of the time factor into the equation requires the solution of a differential equation with complex coefficients. The form of the final differential equation is generally known as the nuclear decay equation. Upon solving this equation, the total P/D Equation of photoexcitation/de-excitation is obtained. **Eq. 16.4.1-2** represents this equation. The equation describes the number of free electrons generated per unit time within the base of the first Activa of the photoreceptor cell following an impulse stimulus. The equivalent response for a square wave stimulus is given in **Appendix.A**. This current profile is defined as a Type A waveform (not to be confused with the *a*-wave of electroretinography) in **Section xxx**. When measured with a voltage probe within the IPM (or within the plasma space of the inner segment associated with the common emitter connection–see **Section 12.8.1**), the signal is designed as a Type B waveform. It is this voltage waveform that is the basis of the *a*-wave of electroretinography.

$$i(F,t,j,\sigma,\tau) := \frac{\sigma\cdot F\cdot\tau}{(1-\sigma\cdot F\cdot\tau)}\cdot e^{j\cdot K_T\cdot\left(\frac{1}{F}\right)^{\frac{1}{6}}}\cdot\left[e^{-\sigma\cdot F\cdot t\cdot K_T}-e^{K_T\cdot\left(\frac{-t}{\tau}\right)}\right]$$

Complete Photodetection/De-excitation equation

Eq. 16.4.1-2

The equation is quite formidable on first viewing. However, it can be broken into several parts. The terms can then be defined in a less cluttered environment. The easiest part to isolate is that between the left bracket and the left parenthesis. This exponential term with an imaginary argument (indicated by the operator, *j*) is the delay term of the equation. It only contains two variables and both the stimulus intensity, F, and the temperature, T, of the animal stand in isolation. This function will de discussed in **Section 16.4.1.1.1**. The remainder of the equation involves real variables. It can be considered the amplitude characteristic of the overall Equation. It will be discussed in **Section 16.4.1.1.2**. Note the presence of two exponential terms within the parentheses. Both contain complicated arguements that are a function of temperature. The argument of the first exponential also includes a unique function of the intensity of stimulation and two other variables, $\sigma\bullet F\bullet\tau$. After discussing and evaluation these fragments, the overall equation will be evaluated, based on the data available in the literature, in **Section 16.4.1.1.3.**

16.4.1.1.1 The scaling factor

The scaling factor of the P/D Equation is another example of nonlinearity in the visual system. For $\sigma\bullet F\bullet\tau$

>>1.00, the scaling factor becomes equal to minus one. However, for values of $\sigma \cdot F \cdot \tau < 1.00$, the scaling factor becomes proportional to $\sigma \cdot F \cdot \tau$.

As developed in **Section 7.2.4**, the absorption cross section, σ, can vary tremendously in vision and is observable via the phenomenon known as bleaching. The product of the absorption cross section and the stimulating flux is maintained as a quasi-constant, $\sigma \bullet F \approx C$, in the human visual modality (as a minimum) by the emitter current of the 1st Activa (the adaptation amplifier) in each sensory receptor neuron. The emitter current mediates the de-excitation of the excitons in the chromophoric coating of the disks of the outer segments of the sensory receptor neurons. As noted below, the time constant, τ , in the human visual modality is typically 125 msec.

16.4.1.1.2 The delay term

The kernal of the complete P/D Equation related to delay is given by **Eq. 16.3.6-2**

$$\Theta(F) := e^{j \cdot K_T \cdot K_F}$$

The delay term of the Photodetection/De-excitation equation Eq. 16.4.1-3

where the delay, $\Theta(F)$, is given in terms of the two auxiliary equations given above. It is the argument of this expression that expresses the delay in units of time. The delay is seen to be hyperbolic with respect to the excitation, F. This fact plays a significant role in the controversy introduced by Cideciyan & Jacobson[40]. **Figure 16.4.1-2** plots the delay in the P/D Equation on both linear and logarithmic scales against the stimulus in absorbed photons/unit area. It shows the delay is changing rapidly for low absorptions and slowly for high absorptions.

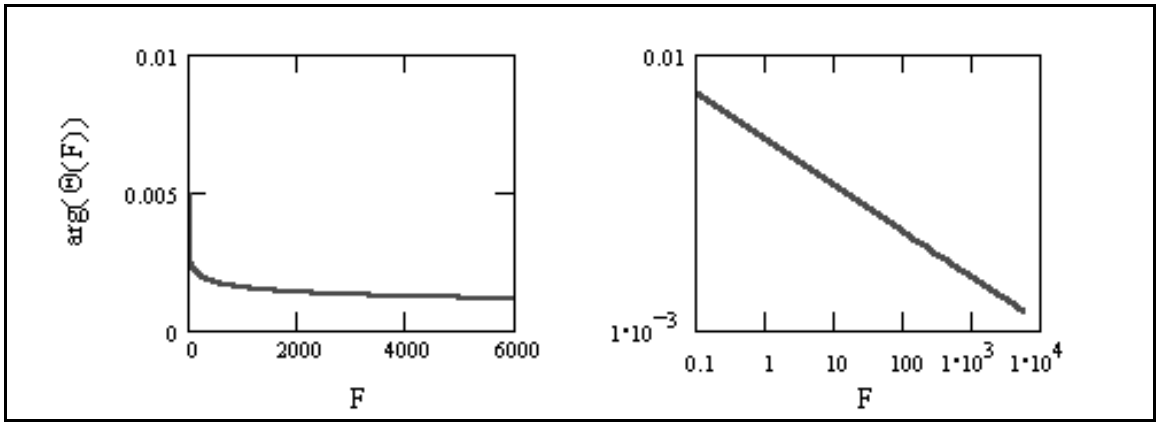


Figure 16.4.1-2 The delay as a function of absorbed irradiance in the P/D Equation. Fd = 1.0, s = 0.01, T= 37 Celsius & t = 12.5 msec.

[40]Cideciyan, A. & Jacobson, S. (1996) An alternative phototransduction model for Human rod and cone ERG *a*-waves. *Vision Res*. vol. 36, no. 16, pp 2609-2621

The P/D Equation reflects several relationships between the excited electrons within the π^* band of the liquid crystal and their environment. It appears that the average velocity of the excited electrons is a function of the number of electrons present and indirectly the illumination intensity. This velocity also appears to be a function of temperature. The delay is inversely related to the average velocity of the excited electrons. As can be seen in the data of Fourtes & Hodgkin, any background illumination present must also be considered in determining the total number of excited electrons at any given time. It is the total number present that determines the average velocity of their propagation to the location of de-excitation in the furrows of the disks

Although the above equation is reminiscent of one used in solid state semiconductor physics, no effort was made to locate the values of the appropriate physical constants for the putative base material. Instead, the literature was searched for relevant empirical data. **Figure 16.4.1-3** illustrates the data located on a grid formed by the above equation (dashed lines). The data generally reflects the generator potential (an analog waveform) at the output of a photoreceptor cell. However, in the case of Limulus, it has been conventional to measure the signal at a point within the eccentric cell. In either case, a very small delay due to the transit time of the signal between the base of the adaptation amplifier and the pedicel of the distribution amplifier has probably been included in this data. Although the data was taken by a variety of investigators over a period of 50 years, it is remarkably consistent. The data conforms quite well to the proposed sixth root of irradiance. All of the data lines were drawn with the highest irradiance normalized to 1.0. While a precise experiment would require the irradiance to be measured in photons absorbed within a narrow spectral range, most of the data was acquired with broad band illumination and measured in terms of photometric units or trolands.

There is an obvious variation in delay time with both irradiance and temperature. **Table 16.4.1-1** summarizes the parameters and sources of the data. This author determined the delay by forming a tangent to the rising portion of each P/D waveform given by these authors and extending the tangent until it

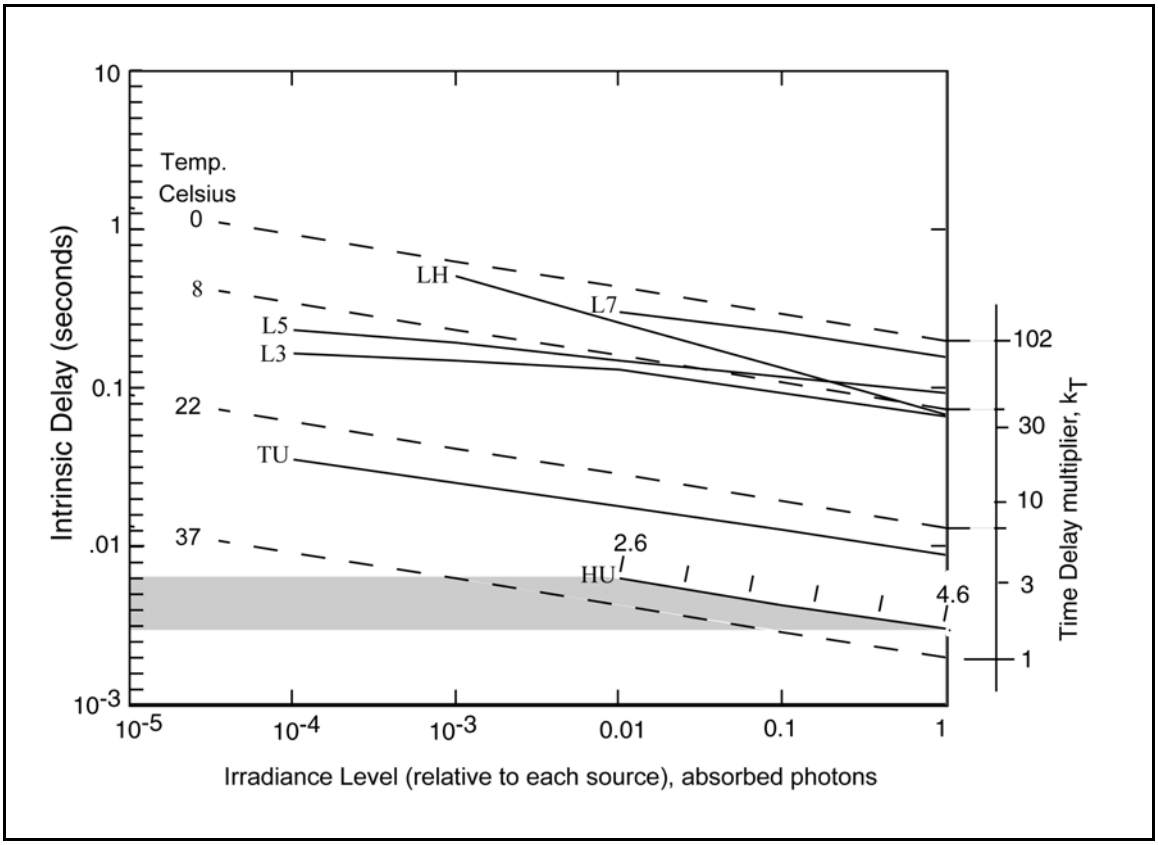


Figure 16.4.1-3 Signal delay in different species as a function of irradiance with temperature as a parameter. L7 is *Limulus* at 7 C (with a background) from Fourtes & Hodgkin. LH is *Limulus* at an unspecified temperature from Hartline. L5 is *Limulus* at 8 C from Fig. 5 of Fourtes & Hodgkin. L3 is *Limulus* at 8 C from Fig 3 of Fourtes & Hodgkin. TU is turtle at 21.8 C from Baylor, Hodgkin & Lamb. HU is human at 37 C from Cideciyan. See text.

intersected the zero illumination baseline. The time interval between this intersection and the time of the illumination impulse was taken as the delay.

The agreement between the data and the temperature grid is less precise. The expression for the temperature shown in the above expression is believed to be unique to biological systems. It relates to a biological temperature range (0 to 46-50 Celsius) that constrains biological activity. The factor of eight used in the denominator of the exponential was derived empirically. This factor determines the temperature range over which biological systems remain viable from a neurological perspective. This range appears to be controlled by the base material of the Activas, a liquid crystal of hydronium. Although requiring further exploratory conformation, it appears that the mobility of the electrons and holes in this material are constrained by the transition of the material into a conventional crystal at a temperature approximating zero degrees Celsius. At temperatures above this level, the mobilities appear to be a function of temperature relative to this transition temperature, not to absolute zero. The relative time constant at a given temperature relative to its value at 37 Celsius is given by the right hand scale of the figure. As an example, an animal at 20 Celsius would expect to exhibit a time constant in the P/D Equation of about 10 times the nominal value at 37 Celsius if the same mechanisms were at work within the P/D mechanisms.

Most investigators did not record the temperature of the eye with sufficient precision. In the case of Cideciyan & Jacobson, the range of delays determined from their graph were within the shaded area. They averaged the data from about 15 human subjects and expressed the stimulus intensity in Log Scotopic Troland-sec. Their average temperature can be expected to be 37 Celsius within a small fraction of a

degree. Since the data of Cideciyan & Jacobson was normalized in this figure, if it had been normalized with respect to 0.1, it would have coincided exactly with the proposed 37 Celsius line.

TABLE 16.4.1-1
Response Delay following impulse irradiation for various species

Curve	Animal	Temp.	Investigator	Comment
L7	*Limulus*	7C	Fourtes & Hodgkin[41]	With background illum.
LH	*Limulus*	unspecified	Hartline[42]	
L5	*Limulus*	8C	Fourtes & Hodgkin	Their figure 5
L3	*Limulus*	8C	Fourtes & Hodgkin	Their figure 3
TU	Turtle	21.8C	Baylor, Hodgkin & Lamb[43]	
HU	Human	37C	Cideciyan[44]	Their figure 1

The only curve with a significantly different slope than the grid lines is that presented by Hartline in the 1930's. He was surely using a mercury-in-glass thermometer. It is important in these measurements that the temperature of the specimen be measured *and recorded* to a precision of better than 0.5 C with equipment calibrated to about one fifth of this value. This is particularly important when working at the lower temperatures. In general, this means the thermal detector needs to be located immediately adjacent to the specimen and not nearby on the substrate or in the cooling bath. In warm blooded animals, the temperature still needs to be measured precisely and recorded.

Unfortunately, only a limited range of data is available for human vision. It is ERG data which also includes time delays related to the travel of the signal from the retina to the test probe at the cornea. Extrapolating the performance predicted by the theory, the total range of delay is over 5:1 in time for an illumination range of about 5:1 in log magnitude. The total predicted range is over 10:1 in time for the range represented by the photopic to scotopic regions.

16.4.1.1.3 The amplitude response of the P/D Equation

The complete amplitude response of the P/D equation is even more complex in form than the delay response. However, it exhibits a series of unique features.

$$i(F,t,j,\sigma,\tau) := \frac{\sigma \cdot F \cdot \tau}{(1 - \sigma \cdot F \cdot \tau)} \cdot \left[e^{-\sigma \cdot F \cdot t \cdot K_T} - e^{K_T \cdot \left(\frac{t}{\tau}\right)} \right]$$

The amplitude function of the complete P/D Equation Eq. 16.4.1-4

The response, i, is a function of a variety of parameters. These include the temperature, the intensity of the input stimulus, F, a time constant for the relaxation of the excitons stimulated by the radiation, the

---

[41]Fourtes, M. & Hodgkin, A. (1964) Op. Cit.

[42]Hartline, (1934) XXX

[43]Baylor, D. Hodgkin, A. & Lamb, T. (1974) Op. Cit.

[44]Cideciyan, A. & Jacobson, S. (1996) An alternative phototransduction model for Human rod and cone ERG *a*-waves. *Vision Res*. vol. 36, no. 16, pp 2609-2621

absorption cross section of the absorbing mechanism, and time. J is the imaginary operator. It plays no role in the amplitude response except as a reminder that there is an associated time delay (presented above).

This equation exhibits a number of important features.

1. It is a first order equation that departs abruptly from the baseline at time zero. The derivative of the equation has a finite value at 0+. This feature is not compatible with the frequently proposed Michaelis equation which is of second order).

2. the exponential expression is a function of temperature as well as stimulus intensity.

3. The equation includes a zero in the denominator under specific conditions. For $\sigma \bullet F \bullet \tau = 1.00$, the equation reduces to a simpler form. This simpler form is the same as Hodgkin's equation (7) for n = 2 (a two stage filter). However, the general form is not compatible with Hodgkin's family of responses. The equation for this special case does offer a unique ability to determine one of the two time constants of the general case.

Figure 16.4.1-4 is a plot of the amplitude portion of the complete P/D Equation for a series of values of the total photon flux per unit area, F, incident on the outer segments of the photoreceptor cells. All of the curves were drawn for a time constant, τ, of 125 msec. The value of σ was 0.01 unit areas. The heaviest line represents the degenerate case, corresponding precisely to the equation of Hodgkin. For this case, the rising and falling time constants for the two slopes of the figure are both equal to 12.5 msec. The absolute delay, where the functions break away from the horizontal axis, are not visible in this figure. They are tabulated, beginning with the highest flux level, in the caption. The graph was prepared using MathCad. 6.0 and the equation as above.

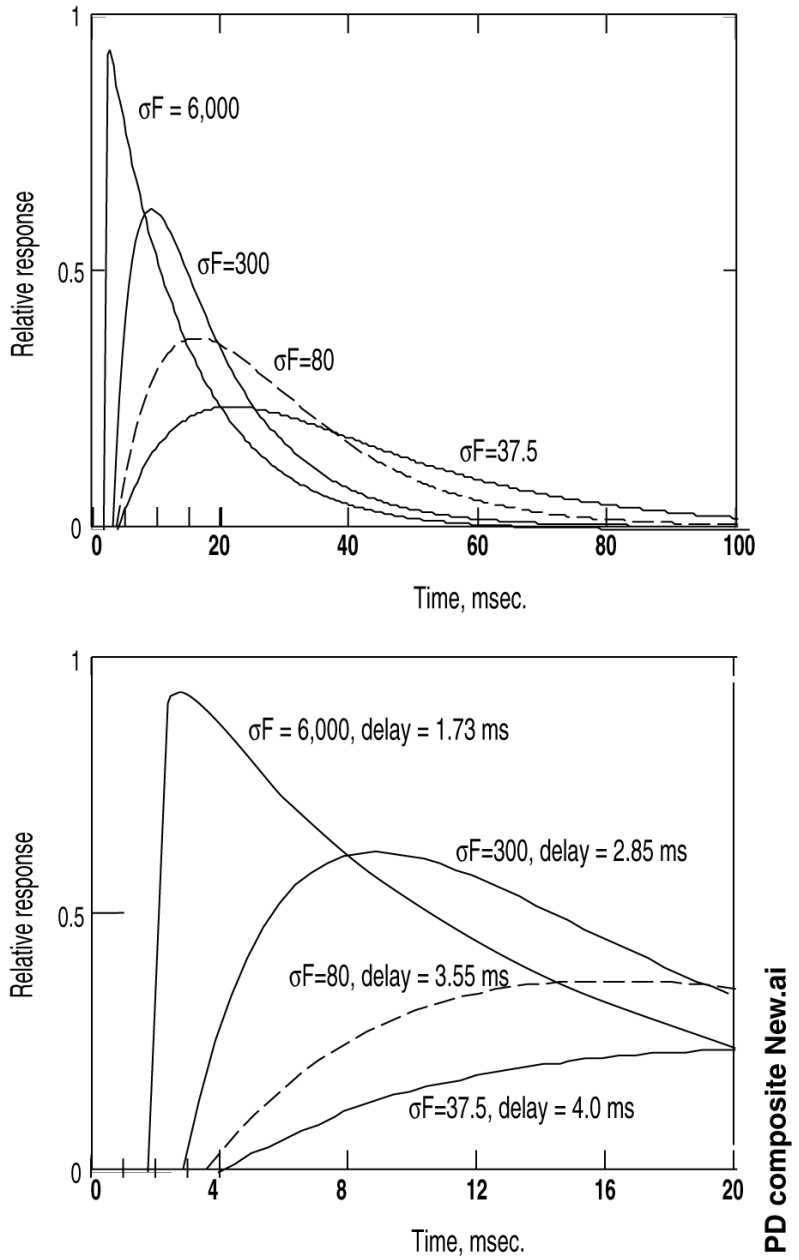


Figure 16.4.1-4 A family of parametric curves for the P/D Equation, shown expanded in the lower frame. Note the abrupt departure from the baseline in the absence of noise. All curves are drawn for a temperature of 37 Celsius. The dashed line represents the degenerate condition, $\sigma \cdot F \cdot \tau$ = 1.00 and corresponds to the Hodgkin solution at 800 photons/unit area-second. The other curves represent the labeled number of absorbed photons/unit area-second, F. The intrinsic delays for each curve are shown in the bottom frame. See text.

Nikonov, et. al. developed an empirical equation containing the difference of two exponentials to describe the impulse response of a portion of the photoexcitation/de-excitation process[45]. However, the terms associated with the arguments were not determined.

16.4.1.2 Evaluating the total P/D Equation under <u>impulse</u> stimulation

16.4.1.2.1 Evaluating the P/D Equation

The total P/D Equation of photoexcitation/de-excitation is quite formidable. Its development is detailed in Appendix A. The equation describes the number of free electrons generated per unit time within the base of the first Activa of the photoreceptor cell following an impulse stimulus. The equivalent response for a square wave stimulus is given in the appendix. The equation consists of three major parts and two subsidiary equations to minimize complexity. The right hand term is the amplitude response of the system. The middle term is the intrinsic delay before the start of the amplitude response. The left hand quotient is a scale factor associated with the equation. The equation contains two coefficients, K_T and K_F defined in the auxiliary equations. The coefficient k_d in the expression for K_F provides the scaling factor needed to account for the variety of units used to describe the intensity of the stimulus and whether the incident radiation is axial or transverse to the outer segments.. For the axial data of Cideciyan & Jacobson collected in terms of log scotopic troland-sec, to be discussed below, k_d is approximately 0.0159. Cideciyan & Jacobson employed a special calculation to determine the "rod-isolated" a-wave from their raw data. This calculation may affect the value of the parameter k_d.

$$i(F,t,j,\sigma,\tau) := \frac{\sigma\cdot F\cdot\tau}{(1-\sigma\cdot F\cdot\tau)}\cdot e^{j\cdot K_T\cdot\left(\frac{1}{F}\right)^{\frac{1}{6}}}\cdot\left[e^{-\sigma\cdot F\cdot t\cdot K_T}-e^{K_T\cdot\left(\frac{-t}{\tau}\right)}\right]$$

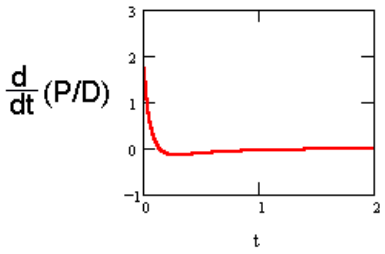
Complete Photodetection/De-excitation equation Eq. 16.4.1-5

The constant value of 102 in the expression for K_T gives K_T the value of 1.000 at 37 Celsius. This allows the time constant, τ, to properly represent the intrinsic time constant of the P/D Equation for humans (*in-vivo* or at normal core temperature). At other temperatures, K_T provides the proper scale factor, where the observed time constant, $\tau_{eff} = \tau/K_T$. In animal experiments, it may not be useful to refer all measured time constants to an intrinsic time constant based on 37 Celsius. By replacing the value 102 by one, the equation will compute all time constants with respect to the intrinsic value at 0 Celsius.

The first derivative of the complete P/D Equation is important to the understanding of the graphs of the Equation. It is given by,

[45]Nikonov, S. Engheta, N. & Pugh, E. (1998) Op. Cit. Eq. 5

$$\frac{d}{dt}(P/D) = \frac{\sigma \cdot F \cdot \tau}{1 - \sigma \cdot F \cdot \tau} \cdot \left[-\left(\frac{\sigma \cdot F}{K_T}\right) \cdot e^{-\left(\frac{\sigma \cdot F}{K_T}\right) \cdot t} + \frac{1}{K_T \cdot \tau} \cdot e^{\left(\frac{-t}{K_T \cdot \tau}\right)} \right] \qquad (14)$$



First Derivative of the P/D Equation Eq. 16.4.1.6

There are important features to be noted in this equation. Many of them depend on the value of the quantity, $\sigma \cdot F \cdot \tau$. As noted earlier, for $\sigma \cdot F \cdot \tau$ equal to 1.000, the basic P/D E equation simplifies to the Poisson Equation for n=2. This form has been labeled the Hodgkin Solution in this work because of the 1964 paper of Fourtes & Hodgkin. The derivative of this simpler form will not be given here.

There is also an important distinction based on whether $\sigma \cdot F \cdot \tau$ is greater or smaller than 1.0. For small values of $\sigma \cdot F \cdot \tau$, the denominator of the scale factor is equal to one. However, for large values of $\sigma \cdot F \cdot \tau$, the value of the scale factor becomes (-1). This negative can be eliminated by removing a factor of (-1) from the bracketed term as well. This causes an important change. The determining factors in the leading and trailing edge slopes change. For $\sigma \cdot F \cdot \tau > 1.00$, the initial slope of the leading edge is given by $1/(\sigma \cdot F \cdot K_T)$ and the initial slope of the trailing edge is given by the time constant τ/K_T. For $\sigma \cdot F \cdot \tau < 1.00$, the initial slope of the leading edge is given by τ/K_T and the initial slope of the trailing edge is given by $1/(\sigma \cdot F \cdot K_T)$. For subsequent times, the slope is changing continually as described by the complete equation. This continual change has prevented many investigators from interpreting the P/D Equation properly.

Many investigators have suggested that the rise time of the Class B, C or D waveforms reach a maximum slope as the stimulation increases. The equation does not support this conclusion. The maximum slope is usually due to mechanisms subsequent to the P/D mechanism. These include saturation in the adaptation amplifier, saturation in the distribution amplifier or the logarithmic function associated with the pedicle.

16.4.1.2.2 Comparing the total P/D Equation to measured impulse data

It is necessary to quantify some of the parameters in this equation in order to apply it properly to the visual process. This can be done best with comparison to actual Class C waveforms. However, there is very little Class C data for humans. There is considerably more Class D data buried in ERG measurements. However, the class D data usually shows considerable compression due to the distribution amplifier of the photoreceptor cells. The above Classes of waveforms were defined in **Section 11.1.5** and illustrated more fully in **Section 12.8**.

Figure 16.4.1-5 Shows the P/D Equation overlaid on the current data of Baylor, Nunn & Schnapf for the monkey, *Macaca fascicularis*[46]. Only the amplitude profiles were provided. No measurement of absolute time delay was made. Each trace is the average of eleven flashes of light. No range bars or precise excitation intensity was given. This data represents the current at the collector terminal of the adaptation amplifier within one photoreceptor cell. For the lower six responses, the P/D Equation overlays the laboratory data precisely. For the upper three traces, saturation has become significant. The level of

[46]Baylor, D. Nunn, B. & Schnapf, J. (1984) The photocurrent, noise and spectral sensitivity of rods of the monkey Macaca fascicularis *J Physiol* vol 357: 575–607.

saturation is indicated by comparing the third highest intensity trace and the predicted response (dashed line). For additional comparisons with this type of data, see **Section 12.8.1**.

The gain of the adaptation amplifier is at maximum for these low stimulus levels. The dynamic range of the adaptation amplifier can be seen from the noise level and saturation level. The ratio of peak signal to RMS noise is between 100:1 and 200:1.

As noted by Baylor, Nunn & Schnapf, the top two traces exhibited a long tail. This may be caused by one of several factors. It may reflect the onset of adaptation where the amplifier gain begins to change with excitation level. The process both lowers the gain of the circuit and begins to establish a new DC level, at the output of the circuit going to the distribution Activa. Alternately, it can be considered an artifact of the adaptation amplifier performance. When forced into saturation, semiconductor amplifiers frequently exhibit a long tail following excitation as they re-establish their normal operating regime. This is due to the motion of charges within the semiconductor required to re-establish the necessary charge density profiles at the atomic level. More analysis should answer this question. See **Section 16.4.1.1.7**.

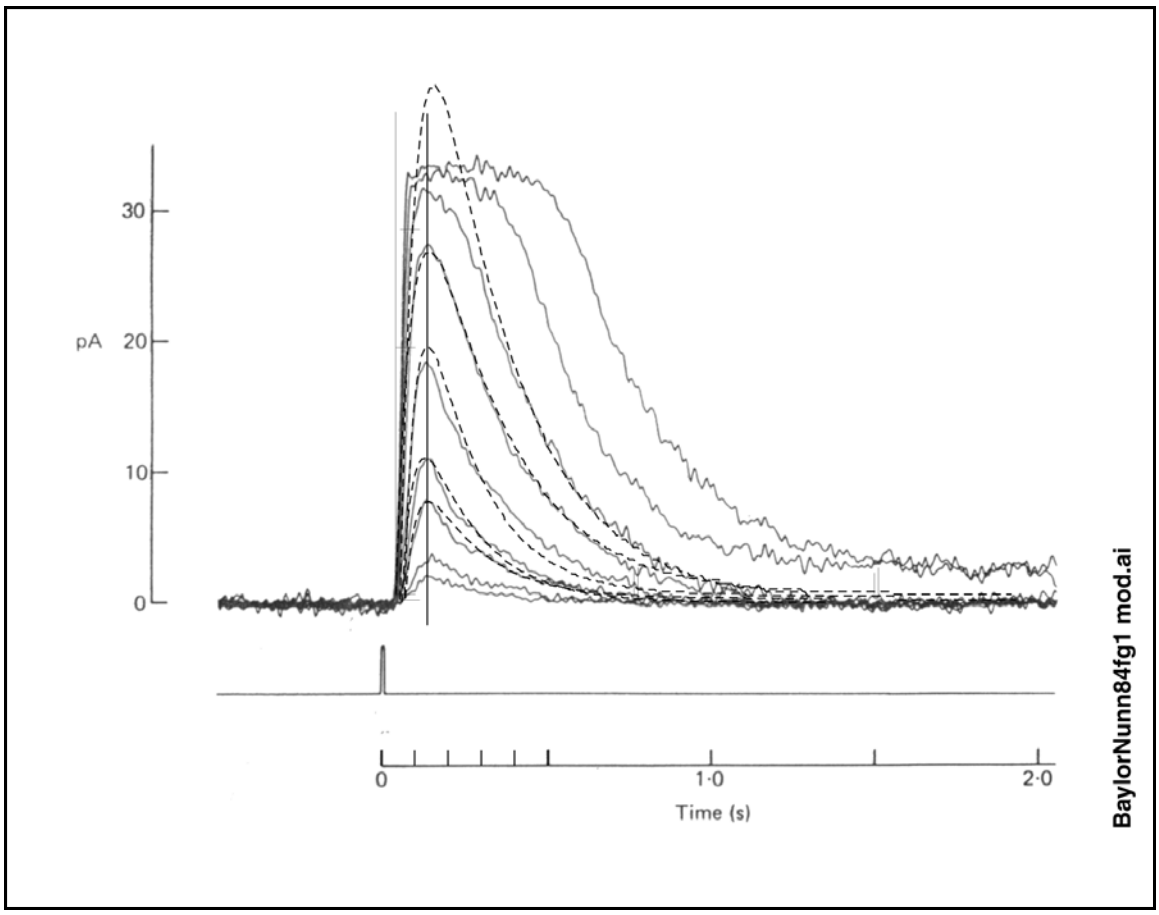


Figure 16.4.1-5 A comparison of the complete P/D Equation versus the data of Baylor, et. al. The solid curves are for a single cell excited by plane polarized transverse 500 nm radiation. The dashed lines show curves based on the P/D Equation. See text.

See **Section 7.2.4** regarding the saturation in these measured responses. The saturation level relates to the maximum current through the 1st Activa (the adaptation amplifier) of the sensory receptor neuron.

Figure 16.4.1-6 compares the complete P/D Equation and the ERG data of Smith & Lamb[47]. The P/D Equation is a demonstrably better fit than the Michaelis curve suggested by Smith & Lamb for the first 15 milliseconds of their measured data labeled 250k Isoms/rod. After 15 milliseconds, the measured curve becomes more complex. The applicability of the P/D Equation to this region will be discussed in **Section 17.6.2**. Note the distinct delay of 3.5 milliseconds before the departure of the P/D Equation from the baseline. The theoretical curve is for a product of $\sigma \bullet F \bullet \tau = 1.875$. The fit shown in this figure used a time constant, t, of 12.5 ms and $\sigma \bullet F$ product of 150 photon-seconds. The time constant appears to match the descending part of the response well. The fit appears excellent. From these numbers, $\sigma \bullet F = 150$ and the flux can be calculated as $1.5 \times 10^{+6}$ absorbed photons per unit area.

When fitting the P/D Equation to ERG data, it is necessary to pick data waveforms that do not show excessive compression due to the amplitude of the stimulus applied. The data of Breton & Schuler, as an example, shows considerable compression. This compression distorts the portion of the waveform due to the underlying Class C waveform.

The best method of calibrating the P/D Equation is using actual Class C waveform data obtained from probes (LERG waveforms, preferably obtained from within the IPM as in the previous discussion).

There is a problem in that the effective absorption cross section of a multi-layered disk stack is not known. If we take the absorption cross section as equal to the physical cross section, $\sigma = 12.5$ sq. microns or 12.5×10^{-12} square meters, the absorbed flux per photoreceptor would be about 12 photons per sq. micron or 150 photons per second per photoreceptor. For a xxx

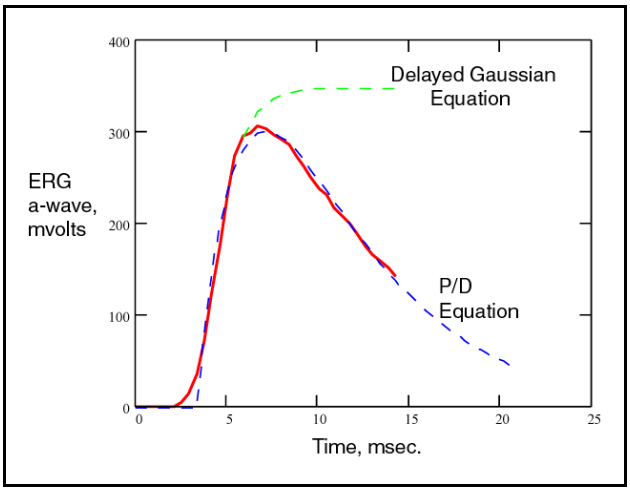


Figure 16.4.1-6 A comparison of the complete P/D Equation versus the ERG data of Smith & Lamb (for impulse excitation). Also shown is the delayed Gaussian equation proposed by Lamb. The solid curve is the measured data for a putative rod-isolated *a*-wave labeled 250k Isoms/rod. The P/D Equation is for $\sigma \bullet F \bullet \tau = 1.875$, τ = 12.5 msec. and $\sigma \bullet F$ = 150 absorbed photon-seconds. The temperature is taken as 37 Celsius.

If an alternate cross-section is used based on the area of a typical outer segment. Cideciyan & Jacobson have recently criticized the formula used in Smith & Lamb[48]. They note the Smith & Lamb formulation exhibits a constant latency and an increasing slope with excitation. They proposed an alternate formula, employing a three-stage filter that exhibits a constant slope at high excitation and a shortening latency. They did not address the low excitation condition. Like Smith & Lamb, their model relies upon three axioms of the conventional wisdom and remains applicable only to the leading edge of the photoexcitation-de-excitation process. Neither of these teams attempted to interpret the fundamental mechanisms underlying the transduction process. The P/D Equation of this work provides the total solution to the process over any range of stimulus intensity and waveshape. The P/D Equation provides a complete explanation for the phototransduction problem. It exhibits both a variable delay and a variable slope of the leading edge as a function of stimulus intensity– in agreement with the measured data.

16.4.1.2.2 The potential measured within the IPM–Class C waveform

---

[47]Smith, N. & Lamb, T. (1997) The *a*-wave of the human electroretingram recorded with a minimally invasive technique. *Vision Res*. vol. 37, no. 21, pp 2943-2952

[48]Cideciyan, A. & Jacobson, S. (1996) An alternative phototransduction model for human rod and cone ERG *a*-waves: *Vision Res*. vol 36, no 16, pp 2609-2621

It is practical to capture all of the current from a single outer segment using the suction pipette technique and to pass it through a resistance to create a voltage. The resulting voltage is defined as a class C waveform.

The actual voltage in the volume of the IPM between the outer segments can be measured. It is the potential caused by the collective currents of the dendrites of the outer segments passing through the impedance of the IPM. This impedance is expected to be resistive. The currents are greater than the currents generated by the P/D equation within the base regions of the adaptation amplifiers by the current gain of those amplifiers. While this current may be large, it is a function of the state of adaptation of the amplifiers. It is also highly localized. Most of this current returns to the individual adaptation amplifiers through the external common emitter contact near the OLM.

It is highly unlikely that an investigator can probe only one dendrite within only one outer segment of the IPM and obtain the voltage profile of that dendrite.

16.4.1.2.3 The potential measured at the pedicle–Class D waveform

The so-called generator waveform is the voltage associated with the pedicle of an individual photoreceptor cell. It is the result of passing the current from the distribution amplifier, a mirror image of the signal current passing through the adaptation amplifier plus a pedestal, through the diode impedance associated with the pedicle. Under small signal conditions, this potential appears to be identical to the current forming it, since the diode is piecewise linear. However, under large signal conditions, the nonlinearity of the diode must be considered. This factor accounts for the compression in the depolarization of the pedicle voltage commonly observed at high excitations.

16.4.1.3 Evaluating the total P/D Equation under finite pulse stimulation

---

[xxx fill in rectangular pulse response]

In summary, the end of the observed plateau in the P/D Equation response does not occur synchronously with the termination of the stimulus. There is a finite delay before the response begins to decay to the baseline. Furthermore, if an incremental change in the illumination follows the initial pulse (such as in the case of pre-adaptation), the P/D response begins with a finite number of excited electrons in the chromophores of the outer segment. The presence of these electrons impacts the shape of the subsequent response function. This fact is illustrated in the next section.

16.4.1.4 Evaluating the total P/D Equation with a background present

Kaplan, et. al. have presented extensive data on the performance of the photoreceptors of Limulus during the circadian cycle of day to night[49]. The peak signal amplitudes of 14 mV make it difficult to determine whether their specific waveforms were of Class C or Class D (See below). All tests used a 10 ms impulse generally derived from a green LED of unspecified wavelength and were conducted at 20 Celsius out of doors. In some cases, an incandescent source was used. They propose that the difference in the shape of the waveforms are due to a change in cell conditions in response to an efferent instruction received from the animals brain. The signal was presumed to exhibit a circadian rhythm. A more direct answer is to associated the change in waveform shape with the background illumination level. Although the mathematics is straightforward, it does not warrant inclusion here. The data can be analyzed based on the above discussion concerning a finite pulse applied to the P/D mechanism. The background illumination causes a level of excited electrons to exist in the π^* band of the liquid crystal chromophore prior to the imposition of the impulse. This is quantified in the caption to their Figure 1. They say the quiescent voltage level of the signal varied from 54 mV during the day to 50 mV during the night. This would suggest that the waveforms were measured in the axoplasm of the photoreceptor cells and were therefore Class D voltage waveforms. The 54 mV would correspond to the normal quiescent hyperpolarization related to photopic operation (**Section 12.5.3 xxx**). The total response is due to the sum of the responses due to a background plus an impulse.

---

[49]Kaplan, E. Barlow, R. Renninger, G. & Purpura, K. (1990) Circadian rhythms in *Limulus* photoreceptors. *J. Gen Physiol*. vol. 96, pp 665-685

The measured waveforms of Figure 2 of Kaplan, et. al. (plotted on an absolute scale) are virtually identical to those in [**Figure 16.4.1-4**] except the time scale reflects the temperature of 20 Celsius. They did not record or display the intrinsic delay associated with the different waveforms. It is probably that the faster and higher responses occurred during the day. They noted that during the night, the intrinsic rise time of the waveforms remained constant but that the decay time constant varied systematically. This is indicative of the $\sigma \bullet F \bullet \tau < 1.00$ condition in the P/D Equation. Under this condition, the falling time constant is indicative of the total irradiance applied to the photoreceptor cell. The lack of any saturation in the waveforms also suggests this low level of $\sigma \bullet F \bullet \tau$ (probably below $\sigma \bullet F \bullet \tau = 10$ during the day and less than 1.0 at night). The average time to peak response was nominally 51 ms at 20 Celsius. A few of their waveforms show a variation from the average time to peak response. Usually the time to peak tends to decrease with higher response (during daytime).

A cursory application of the P/D Equation to their data suggests the time constant of the P/D Equation for this animal at this temperature is 80-125 ms. This would equate to a time constant of between 8 and 12.5 ms at 37 Celsius. The similarity to the nominal time constant in man of 12.5 ms suggests the same mechanisms are at work within the *Limulus* outer segment as in man.

Kaplan state that the area under each of their responses is 2.4 times greater at night than during the day. This conclusion appears to have been drawn from their normalized data in Figure 2. The absolute data in Figure 2 appears to tell a different story. The total areas appear to be about equal. This would be the expected result from an impulse stimulus. They also note that their data cannot be satisfied using the Poisson Equation of arbitrary degree using a single time constant in multiple stages. Their conclusion that more elaborate multistage models are required overlooks the possibility of a simpler solution such as the P/D Equation. Interestingly, they noted the change in "photoreceptor gain" with increase excitation at night. This increase is precisely that predicted by the scale factor of the P/D Equation as the values of $\sigma \bullet F \bullet \tau$ rise toward 1.00.

Kaplan et. al. conclusively demonstrate the circadian character of their data. They propose that the circadian rhythm of their recordings was due to an endogenous source and they assign this source to an efferent neural signal from a clock located in the brain. They suggest this efferent signal causes a delay in the decay process at night. While this is a rational supposition in the absence of more complete model of the process and the deeper analysis the model would provide, it does not appear to be the actual situation. This work shows that the endogenous source is the level of excited electrons in the liquid crystalline chromophore of the outer segments. This source is in fact due to an external source, the background illumination level falling on the outer segments of the photoreceptors. This background varies with the time of day with a circadian pattern.

16.4.1.5 An extended P/D Equation

More important results emerge from the comparison, particularly with regard to **Figure 16.4.1-5**. Although not noted by Baylor, et. al., all of the non-saturated traces peak at the same time after excitation. Furthermore, all of the theoretical traces were for the condition, $\sigma \bullet F = 6$, $\tau = 0.035$ seconds. The product of these terms gives $\sigma \bullet F \bullet \tau = 0.21$. This value is far below the degenerate condition. In this region, τ represents the time constant of the leading edge and $1/(\sigma \bullet F)$ represents the time constant of the trailing edge. Thus all of the traces exhibit the same leading edge time constant of 35 milliseconds. Their apparent difference in slope is due to the different scale for the traces due to more intense excitation. They all reach a peak after 150 ms according to the investigators (the published artwork suggests a value of 135 ms). A similar condition is described in **Section 12.8.1.2**. Taken together, the data suggests that that the velocity of excited electrons in the π^* band has been reduced. The number of electrons in a given region is so low that they are no longer subject to the enhanced velocity associated with their mutual attraction. It can be surmised that the electrons are not subject to the forces of mutual repulsion. They may be only responding to thermal considerations (Brownian motion). Therefore, the transport equation used to determine the differential form of the P/D Equation needs to be modified. The effect is to insert a term such as $(\sigma \bullet F + 6)$ in place of $\sigma \bullet F$ in the second exponential and the denominator of the scaling term of the complete solution. The transfer function of the P/D mechanism becomes an essentially linear function of the excitation in this region. Further analyses are needed to define the specific form of this second order P/D Equation.

16.4.2 Temporal equation of <u>signal adaptation</u> in stage 1

No theoretical development of the equations of adaptation were found in the literature. Historically, the anecdotal statement that the dark adaptation characteristic is indicative of the transition from cone to rod vision is common. Alternately, Hecht, et. al. offered the suggestion that the characteristic was due to two different mechanisms involved in the reconstitution of visual purple. The later is more nearly in consonance with the following analysis. However, the two mechanisms are both neurological rather than photochemical. Frumkes has recently collected a large amount of the graphical literature related to adaptation[50].

[xxx edit]
Adaptation is characterized by variations in the gain of the adaptation amplifiers within the photoreceptor cells. As a consequence, the values of C_x' become variables as a function of input intensity on a spectral channel basis. This introduces a fundamental change in the matrix of [Eq. xxx] and the equivalent expression for just the luminance channel (equation 3) in Wandell[51]. It is this change that introduces the phenomenon of color constancy into biological vision.

Aside from the very rapid processes involved in signal generation discussed above, the temporal aspects of signal processing in vision are centered in only three areas; the operation of the iris, the adaptation amplifier and the codec circuits. The remaining circuits provide a wider bandwidth than required for undistorted signal transmission. The codec circuits will be discussed in the next section.

As indicated in an earlier section, the closing of the iris is only indirectly related to the perceived brightness as a function of time. Its principle purpose is to improve the resolution performance of the lens/cornea with increasing light level (a function related to the Stiles-Crawford Effect). The change in signal level related to changes in the diameter of the iris are minuscule (about sixteen to one) compared to the changes associated with the adaptation amplifier (typically greater than 3400 to one).

The complete adaptation equation is dominated by the role of the adaptation amplifier. The underlying processes are complex. The attack portion of the adaptation function, relating to light adaptation, is a function of the illumination level applied to the visual system as well as the state variables present in the adaptation amplifier at the time. The decay portion of the function, occurring on termination of any input illumination to the eye is simpler.

The adaptation characteristic usually presented for the human eye are in terms of the minimum signal modulation required to exceed an undefined threshold and thereby be perceived by the subject. The characteristic is usually defined only for the time after cessation of some level of background illumination, usually unspecified but presumably in the photopic range. The object of this Section is to present the entire adaptation characteristic as a function of time with the background illumination level as a parameter.

Figure 16.4.2-1 is a simplified version of the Overall Schematic Diagram of vision used earlier. It shows only those elements of the system related to adaptation. The role of the vascular system will be discussed in detail below. Note also the presence of the closed loop servo system controlling the iris. The signal path of this loop is only shown conceptually. No exhaustive literature search was made to determine the signal manipulation between the photoreceptor cells and the portion of the tectum controlling the muscles of the iris. In the general case, a pretectum may not be present. It is assumed that signals from more than one chromophoric channel is used in the normal operation of the eye. However, the literature is sparse in definitive experiments in this regard. There may even be separate photoreceptors, not related to the perceptual process, supporting the operation of the iris servo loop. It is also assumed that the signals supporting this function are encoded by the projection system in order to reach the brain from the retina. Based on the literature, the gain of this servo loop is relatively low. In human, the optical transmission of the pupil only varies by a factor of 16:1 in response to a change of illumination of several orders of magnitude. This fact supports the conclusion that the iris is primarily a device for controlling the quality of the optic ray bundle passed through the lens group.

---

[50]Frumkes, T. (1990) Classical and modern psychophysical studies of dark and light adaptation, *Chapter 7 in* Leibovic, K. *ed*. Science of Vision, NY: Springer-Verlag pp 172-193

[51]Wandell, B. (1989) Color constancy and the natural image. *Physica Scripta*, vol. 39, pp 187-192 *also in* Byrne & Hilbert, 1997, pg 168

16.4.2.1 The equations of adaptation

16.4.2.1.1 The conceptual equation of adaptation EMPTY

There are three primary components of the adaptation process in vision. xxx

16.4.2.1.2 The electro-vascular characteristics of adaptation

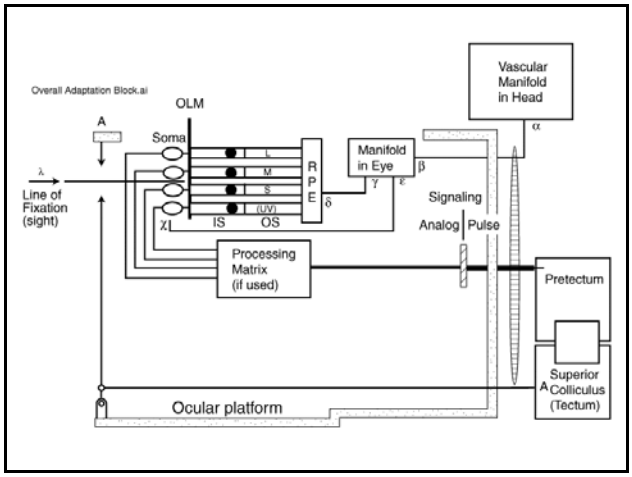


Figure 16.4.2-1 Overall adaptation block diagram including the vascular energy supply sub-system and the iris control servo sub-system.

Although generally not discussed in the literature, the vascular sub-system of the eye plays a crucial role in the transient performance of the eye. This sub-system is the only source of energy for fueling the operation of the eye. The above figure shows several reservoirs within the vascular sub-system. These were discussed briefly in **Section 3.3.1.** The primary reservoir is the richly supplied vascular system of the brain itself. After traveling down the optic nerve, the vascular system of the optic globe divides first into two branches serving the anterior and posterior portions of the retina (separated by the Outer Limiting Membrane, OLM). The branch serving the posterior face divides again to serve the sclera and musculatura of the eye, and the RPE via the choroid artery and veins. It is the branch serving the RPE that is a major factor in the transient operation of the signaling capabilities of the eye. This branch also serves the Outer Segments of the photoreceptors by supplying the Inter-photoreceptor matrix, IPM. As noted in the literature, the cells of the RPE are closely packed and it is required that any nutrients destined for the IPM pass through the RPE cells. Thus, the electrostenolytic processes found on the the surface of the individual photoreceptor neurons depend on nutrients that have passed from the blood stream through an ever more restrictive vascular path. To evaluate this situation, it is useful to employ conventional simulation techniques and define an electrical analog of the vascular system. Such an analog is shown combined with the actual electronic adaptation circuit of the photoreceptor cell in **Figure 16.4.2-2.** The reservoirs of the vascular system are indicated by the capacitors underneath the voltages, V_1 through V_4 in the figure. The impedances of the vascular sub-system are represented by the resistors shown between these voltage nodes. The RPE and photoreceptor cell walls are shown by the dotted boxes. The other resistive elements are representative of the impedance related to fluid flow in the system. This flow may be limited to diffusion in the vascular bed represented by the reservoir at V_3 and V_1. The resistor symbol shown surrounding most of the photoreceptor cell and terminating at the OLM is meant to represent the impedance of the IPM to electrical current flow. The dotted line between this symbol and the OLM is meant to represent the ground bridge between the IPM and the INM. This element is not relevant to this discussion. The ground symbols shown in the figure are parts of the vascular system analog and are not connected to the IPM.

To simplify later discussion, the two electrostenolytic processes on the surface of the dendrolemma are shown differently. The process shown connecting the emitter of the Activa to the IPM is shown as a voltage source and diode in series. The electrostenolytic process, represented by the semi-circular element on the surface of the dendrolemma of the photoreceptor cell, is shown as an impedance allowing current to flow from the voltage node V_1 to the voltage node V. This impedance related to the electrostenolytic element forms the transition between the vascular analog on the left and the actual electrical circuit of the photoreceptor cell on the right. Whereas the other capacitors in the figure are analogs of reservoirs, the capacitor labeled C_A represents the real total capacitance of the dendrolemma associated with the distributed Activa labeled AT. This Activa is shown on the surface of the dendrolemma. It should be shown bridging the space between the IPM and the dendroplasm. At this location, the dendrolemma and the internal reticulum are separated by a minuscule distance filled with hydronium. The collector of the Activa is actually formed by the wall of the reticulum. The thinness of the combined dendrolemma and reticulum results in a large capacitance per unit area between the IPM and the dendroplasm. It is important to note that the Voltages, V_1 through V_4 are nodes in the analog that are shared with an ever wider population of photoreceptor cells (and other consumers of the electrostenolytic energy sources). V_3 serves the entire retina. V_4 serves the entire eye.

Although addressed entirely from an electrical perspective related to the ERG, Rodieck has provided a figure that can be related to the vascular system developed here, in the last and following figure, through the addition of reservoirs between the resistive elements[52].

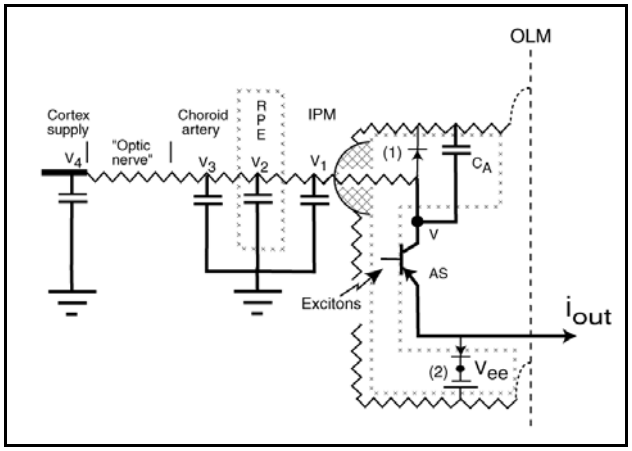


Figure 16.4.2-2 Expanded PC-RPE interface showing electrical analog of the vascular system and the electrostenolytic process supporting the Type AT Activa.

16.4.2.1.3 The operation of the adaptation amplifier

The operation of the adaptation amplifier is a unique feature of the photoreceptor cell in the animal vision system. It shares its unique features with a similar unique amplifier in solid state electronics. The amplifier employs a high degree of negative internal feedback. This feedback factor is so high that the emitter current of the circuit would be constant in the absence of other circuit elements. The critical circuit element is the capacitor C_A. This element, in conjunction with the impedance of the diode and the electrostenolytic source, exhibits a time constant (of about 0.1 Hz in human). These circuit elements constitute a low pass filter in a feedback circuit. The low pass circuit appears as a high pass circuit in the signal path where it defeats the negative feedback mechanism at frequencies above the 0.1 Hz value. The result is two fold. The *average* output current of the circuit will be stabilized for any input excitation above a set value by an adjustment in the current gain of the circuit. The same adjusted current gain is used to amplify the *dynamic* signal within the passband above 0.1Hz.

> In designing vision experiments, it is important to take note of this time constant. Many past experiments using a flickering illumination source have encountered difficulty because their data included the effects of this unrecognized time constant.

Normally, the voltage at the collector of a semiconductor device, such as the Activa, does not impact the operation of the Activa. However, because of the sensitivity of the avalanche effect to the voltage between this voltage and the voltage of the base terminal, the amplitude of the voltage on the node labeled V is critical to the operation of the adaptation amplifier. This voltage can be determined by calculating the sum of the currents entering this node. In the absence of illumination, no current will flow through the Activa. The steady state voltage of the node will be determined by the current flowing to the node from the voltage V1 and away from the node to the IPM through the voltage V_{CC} and the diode representing the dendrolemma. With illumination, current will flow through the Activa, thereby reducing the voltage, V, to a new steady state value. During the transient phases leading to new steady state values of V, the current to or from the capacitor, C_A, must also be considered. Since the voltage between the emitter and the base of an Activa remains nearly constant, the electrical field within the collector of the Activa is determined primarily by the voltage between the node, V, and the voltage at the emitter. For zero current through the Activa, this voltage is essentially the intrinsic dendroplasm voltage minus the voltage V_e. The voltage V_e is essentially constant because of the contribution to the current through the diode associated with V_{ee} from the other Activa in the differential pair circuit of the photoreceptor cell. Therefore, under normal operating conditions, the voltage, V_{cc} minus V_e determines the electrical field within the collector of the adaptation Activa. The amplification factor, or gain, of the Activa is known to be given by this voltage raised to a power. The power is normally between 2 and 6 in other semiconductor devices. The actual values of the electrical elements in this combined analog and real electrical circuit have not been determined. However, both the ON and OFF transient characteristics are known (See Section 17.6.1 for the values for human). These waveforms provide an approximate value for the most important of the circuit elements.

---

[52]Rodieck, R. (1969) The vertebrate retina. San Francisco, CA: W. H. Freeman

16.4.2.1.4 The equations of the adaptation amplifier–the Exposine function

[xxx edit figures to better represent the adaptation amplifier as currently understood]

The adaptation amplifier of the photoreceptor cell exhibits the most dynamic operation of any known neuron. **Figure 16.4.2-3** portrays the above schematic in standard block diagram form suitable for describing both the DC signal performance and the AC signal performance of the amplifier. Note that there is no external feedback path in the actual circuit. The feedback is internal. The RC circuit in the feedback path controls the operation of the circuit. It has a major impact on the slowly varying output of the circuit and no impact on the rapidly varying output. The time constant of the circuit is nominally 33 msec. The transfer function for any variation in the input with a time constant longer than 33 ms. is suppressed according to the equation

$$A = \frac{1}{1 - G} \qquad \text{Eq.16.4.2-1}$$

where G is nominally xxx. For any input component having a time constant significantly shorter than 33 ms., the transfer function is obtained without regard to the feedback loop. A is approximately equal to the forward gain of the amplifier, G.

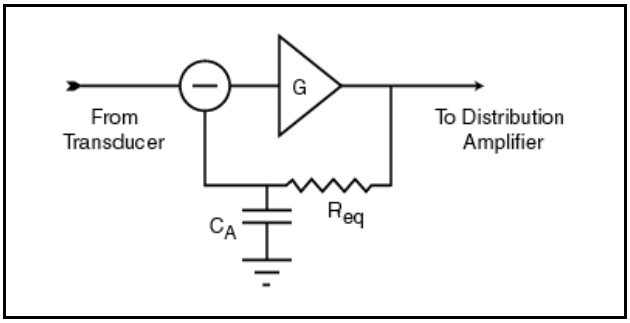


Figure 16.4.2-3 Adaptation amplifier block diagram. The amplifier with gain, G, is in the forward signal path.

Figure 16.4.2-4 defines the idealized circuit schematic in the context of a combined electronic and hydraulic system supporting only a single adaptation amplifier. The system consists of a multistage RC network formed by the vascular and electrostenolytic elements supporting a given photoreceptor. This network removes charge from the capacitor, C_A. The current generator, I, represents the current flowing through the Activa in response to a constant input irradiation flux. Although not required, a switch is shown to define the stoppage of current from the transducer entering the Activa. With no current into it at the emitter, no current is delivered by the collector of the Activa to the capacitor, C_A.

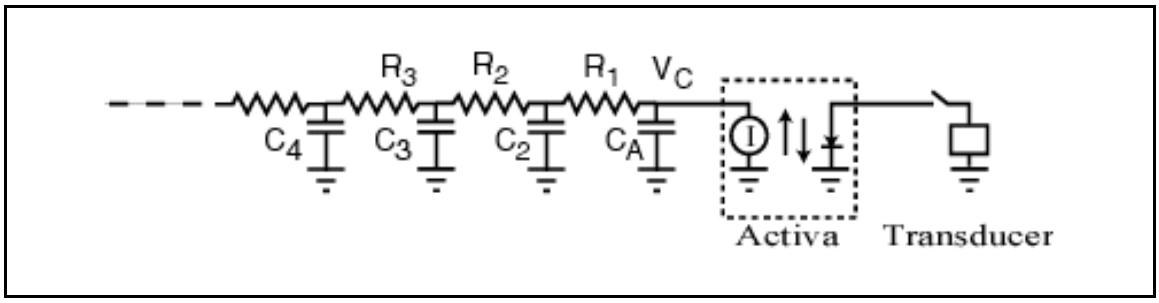


Figure 16.4.2-4 Idealized circuit schematic representing the vascular, electrostenolytic and electronic elements of the adaptation amplifier collector circuit. Opening the switch causes the current, I, to go to zero instantly in this example. In the general case, the voltages on C_2 & C_3 are shared with other adaptation amplifiers.

At this time, it is not clear how many stages of the RC filter network are important to the operation of the individual photoreceptor. **Section 3.3.2** discussed the impact of this network on groups of photoreceptors and the ultimate impact on the perception of edges in the field of view. Notation is adopted in this figure

that suggests the order of the system as a function of the number of RC stages. The network consisting of R_1 and C_A alone would represent a first order system. By adding R_2 and C_2, the overall system becomes second order, etc. The detailed mathematics associated with these networks becomes difficult after the second order system. However, the third and fourth order systems can be represented by the product of various combinations of the first and second order systems. In the operation of the adaptation amplifier documented in the literature to date, it appears only a second order system need be considered in most cases. There is some indication that the photoreceptors of the fovea are supplied by a network in which the equivalent capacitor C_2 is very small. As a result, the amplitude of the sinusoidal component defined below may be very small or the time constant related to R_3C_3 may be exhibited more prominently.

The standard form for the equation representing a second order physical system is given by:

$$\frac{\ddot{q}}{\omega_n^2} + \frac{2\zeta}{\omega_n}\dot{q} + q = q_f \qquad \text{Eq. 16.4.2-2}$$

The complete solution of this equation is written in one of three forms depending on the value of zeta. Each complete solution is usually written in two separate parts representing the transient solution and the "forced solution." The transient solution alone applies to the dark adaptation situation under the assumption that the current through the Activa ceases abruptly (neglecting the transient associated with the P/D equation). Light adaptation under laboratory conditions can be treated using the forced solution under the assumption that the exciting irradiance is a step function.

The combined vascular and electrostenolytic network appears to involve values of zeta near or less than 1.000. For zeta equal to one, the transient equation is a simple exponential. For zeta less than one, the transient equation contains a simple exponential and a sinusoidal term. For zeta greater than one, the equation contains a simple exponential and a hyperbolic sinusoid (the first time the author ever had to address such a term since college). For values of zeta greater than but near 1.00, the hyperbolic sinusoid resembles the associated exponential. The appearance of the total transient response in this case is not significantly different than in the zeta equal to 1.00 case. It will not be addressed further here. The important case is for zeta less than one. That is the case, when combined with other processes, that defines the dark adaptation characteristic recorded psychophysically. This transient solution of the second order system is given by[53]:

$$q_t = e^{-\zeta\omega_n t}\left(A\cos\omega t + \frac{B}{\sqrt{1-\zeta^2}}\sin\omega t\right) \qquad \zeta<1.00 \qquad \text{Eq. 16.4.2–3}$$

This second portion of the equation can be re-written in the form, C x sin ($\omega\tau + \phi$). The overall equation is that of an exposine function, a product of an exponential function and a sinusoidal function.

For ζ=1.00, the equation reduces to:

$$q_t = e^{-\zeta\omega_n t}\left(A + B\omega_n t\right) \qquad \text{Eq. 16.4.2-4}$$

In treating only an isolated adaptation amplifier, the maximum amplitude of the sinusoidal terms and of the exponential term are the same, unity. However, the introduction of additional current paths emanating from C_2 and C_3 removes this commonality. The initial conditions now allow the amplitude of the sinusoidal terms to vary independently of the exponential term that is still exclusive to the individual adaptation amplifier.

The sharing of the voltage nodes between adaptation amplifiers and the introduction of this additional degree of freedom complicates the experimental situation

---

[53]Trimmer, J. (1959) Response of physical systems. NY: John Wiley & Sons. pp. 49-89

considerably. The collector voltage of a single adaptation amplifier is now determined by the state of the adjacent adaptation amplifiers. This fact is the principle complication hindering the understanding of the typical dark adaptation characteristic taken under various chromatic conditions. Use of a chromatically selective test probe following pre-adaptation to"white" or chromatic light must consider this consequence.

Figure 16.4.2-5 shows the principle features of these equations for the general case. The nomenclature is arbitrary. The left and bottom scales represent the mathematical equations. The right scale represents the collector voltage of the adaptation amplifier to be discussed later. In the case of a simple first order system, the collector voltage is at V_q based on the illumination level and the intrinsic potential of the electrostenolytic process at the surface of the axolemma. When the illumination ceases (the switch is opened), current ceases to flow through the Activa. The collector voltage rises along the exponential curve of the figure to the intrinsic electrostenolytic voltage, V_{int}. In the case of the proposed second order system, the response consists of the product of a simple exponential and a sinuosoidal component known as a damped sinuisoid. The detailed shape of the resultant curve is strongly dependent on the values of the individual circuit element. It also depends on the initial condition of the circuit. If the derivative of the voltage on capacitor C_1 is zero, a zero in the sine wave occurs at zero time. Otherwise, the sine wave exhibits an initial phase shift at time equal to zero. The curve shown is for a circuit with t = 20 and $\omega=\pi/15$ [check these values] in arbitrary units of time. The peak amplitude of the sinusoid is taken as 20% of the peak amplitude of the exponential. Note the zero crossings of the sinusoid corresponds to points where the overall waveform has the same amplitude as the exponential component. The peaks in the sinusoid correspond to horizontal points in the right half of the overall waveform. The two arrows suggest how the shape of the damped sinusoid changes as the amplitude of the sinusoid increases relative to the exponential. Notice the arrows introduce opposite polarity changes for the same increase in amplitude of the sinusoid. These features will become important in later sections and the next Chapter.

[The horizontal components of the change predicted here were observed and noted in the data of Hecht, et. al. in 1937. At that time they gave no explanation for the effect. They indicated clearly that the effect changed daily. They also said there preparation was limited to "a 10-15 minute stay in the dim light of the dark room in order to remove the effects of the variable conditions to which the observer has been exposed." This was followed by a light adaptation " for precisely 2 minutes." The limited preadaptation time suggests that the vascular system related to the equivalents of R2-C2 were not adequately stabilized, resulting in the observed variation.][May move this to Chap 17]

Table 16.3.5-1 provides a quick reference to the features of this waveform.

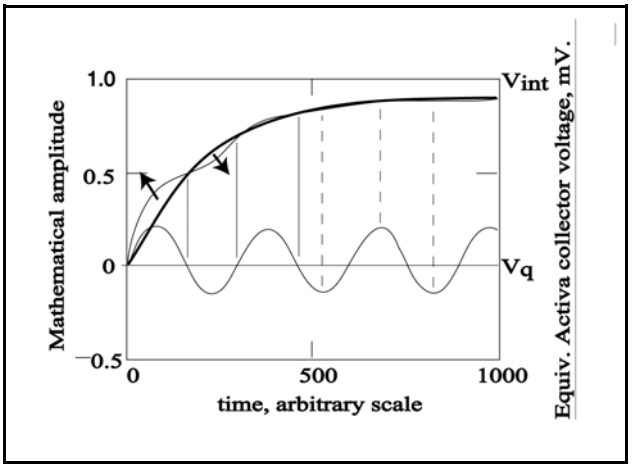


Figure 16.4.2-5 Predicted collector voltage of adaptation amplifier during dark adaptation. Exponential is a first order approximation. Solid line represents proposed second order system. V_q is the quiescent voltage before adaptation. V_{int} is the intrinsic potential of the electrostenolytic process at the axolemma. Time scale is arbitrary.

TABLE 16.3.5-1
FEATURES OF THE COLLECTOR WAVEFORM

1st max. in sine wave	$\pi/2$	
1st zero in sine wave	π	First access to true value of exponential
2nd zero	2π	Second access to true value of exponential
2nd minimum	$7/2\pi$	First definitive appearance of sinusoid
3rd minimum	$9/2\pi$	Second definitive appearance of sinusoid

16.4.2.1.5 Dark adaptation characteristic computation worksheet

Figure 16.4.2-6 displays the complete theoretical derivation of the dark adaptation characteristic. Under low stress conditions, the characteristic consists of just a single inverted exponential function. Under higher stress conditions, the characteristic is represented by an inverted product of an exponential and a damped sinusoidal function multiplied by the exponential (lower graph). The time constants for both functions in the case of human vision are given in minutes. Some of the earlier dark adaptation measurements were taken over time periods of less than a single time constant of the underlying exponential (30 minutes in the case shown. In other experiments, the measurements were taken over intervals of less than the period of the sinusoidal waveform (12 minutes in the case shown. In both cases, the real features of the dark adaptation characteristic can not be interpreted correctly.

For $m > 0.23$ in the above work sheet, the first plateau in the lower blue line will actually turn upwards in the interval between 10 and 12 minutes before descending. This feature has been observed during some laboratory investigations of the dark adaptation characteristic. By appropriate selection of the constants in this set of equations, all three of the data sets provided by Perlman in **Section 18.8.3.6.2** can be matched to any degree desired.

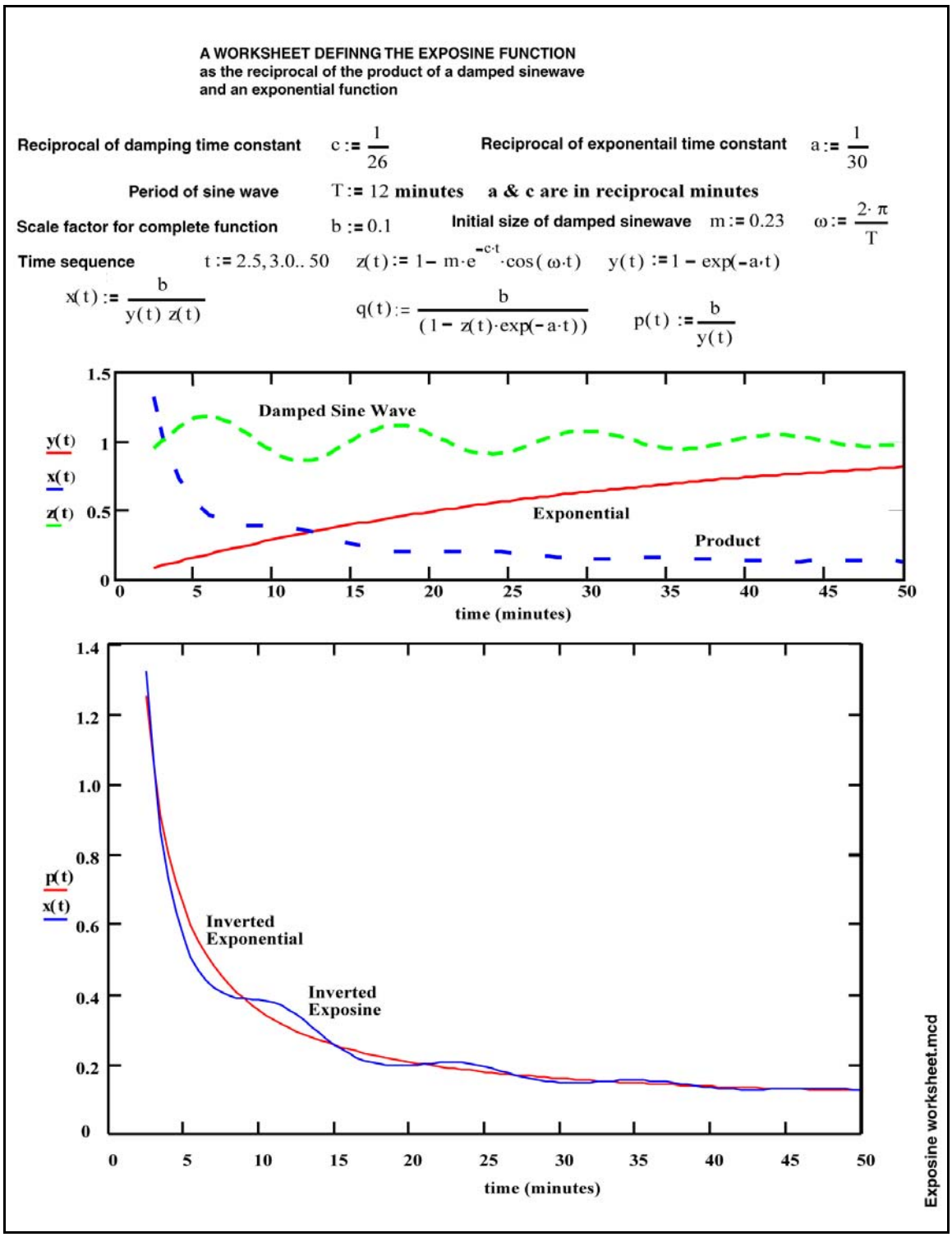


Figure 16.4.2-6 The theoretical dark adaptation worksheet. The equations are those developed using the nomenclature of the program, "Mathcad 6.0." The upper graph shows the two components and their product. The lower graph shows the inverse of the underlying exponential and the inverse of the complete characteristic described here as an exposine function.

16.4.2.2 Comparison of adaptation with the data base of R. E. Records

Brown, writing in Records, appears to have provided the broadest and most recent summary of data related to adaptation[54]. He discusses the "Methods of Measurement of the Dark–Adaptation Process" before proceeding. He investigated the effect on dark adaptation of various intensities and durations of light adaptation prior to the measurement of dark adaptation. He also explored the effect of color on the dark adaptation characteristic. He also collects large amounts of more recent data than that of Hecht and colleagues. Many of his curves fitting the data are monotonic and do not incorporate a discontinuity in the region of 6-15 minutes after the start of adaptation, Figures 14–4, –5, & –6. In other figures, he merely defines two continuous monotonic functions that intersect and cross in this region without asserting any functional continuity between the two regions of data collection involving a discontinuity, Fig. 14–9. In the upper left trace of Fig. 14–9, the exposine character of the data set is obvious although the investigators attempted to fit a simple exponential–like curve to the data set.

Brown makes a remark concerning measurements that do not support the chemical theory of photo transduction and favor the electrolytic theory of this work. After reviewing the "very carefully conducted and thorough experiment on the relationship between the amount of pigment bleaching and duration of the bleaching exposure" of Pugh, and the earlier investigations of Rushton and by Alpern, Brown asserts, "They indicate that the recovery of sensitivity during dark adaptation is not related in some direct way to the recovery of photopigment following bleaching. Rather, Pugh suggests, recovery of sensitivity must relate to some byproduct of the initial molecular absorption of quanta which is outside of the chain of photochemical recovery events in the retina." As shown in this work, the mechanism(s) involved are quantum mechanical and not chemical in character (**Chapter 5**). No chemical byproducts are generated or reconstituted.

Browns collection of data, from many papers in the literature, is very compatible with the theoretical, and monotonic and continuous, exposine function of this work.

16.4.2.2.1 The nominal characteristics for dark adaptation EMPTY

[this section is obsolete, see above definition of expanded damped sinusoid]

16.4.2.2.2 Evaluating the parameters in dark adaptation EDIT

[this section is obsolete, see above definition of expanded damped sinusoid. The caption has already been modified to represent the exposine function.]

Although many dark adaptation characteristics appear in the literature, usually based on the work of Hecht, et. al., they are frequently inadequately documented. **Figure 16.4.2-7** is reproduced from one of their original works[55]. Hecht, et. al. carried out an extensive test program to understand the different situations that might occur in adaptation. Unfortunately, by today's standards, their instrumentation was primitive. They describe their adapting light as "whole (white) light" and indicate their light source was a flashlight bulb operating at 3.8 volts and 0.28 amperes. It would be expected that the color temperature of this bulb is near 2400 K. Although later references to these authors frequently refer to the two branches of the typical adaptation characteristic in humans as related to two separate photoreceptor types, the "rods" and "cones," these authors also introduced another regime. Quoting from their 1937 paper; "There are two types of rod

---

[54]Brown, J. (1979) Adaptation *in* Records, R. *ed*. Physiology of the Human Eye and Visual System NY: Harper & Row, chapter 14, pp 351-389

[55]Hecht, S. Haig, C. & Chase, A. (1937) The influence of light adaptation on subsequent dark adaptation of the eye. J. Gen. Physiol. Vol. 20, pp. 831-850

dark adaptation, a rapid and a delayed, which are probably the manifestations of two methods of formation of visual purple in the rods after its bleaching by the pre-adaptation lights." There work was reported in a series of extensive papers. In the 1937 paper, they detail five special points regarding their test arrangements. One of the most important is that their measurements do not represent the threshold of visibility but a brightness approximately three times higher, based on the appearance of a black cross as criterion. They did not discuss how they determined the factor of three. A precision error of 300% is a large factor in modern laboratory measurements. They also describe several piecewise adjustments to bring their data into better conformity (pg. 841). They used a violet filter (Corning #511transmitting below 480 nm.) and a red filter (Wratten #88 transmitting beyond 690 nm.). They expected the violet light to provide the greatest and the red light to provide the least definition between cone and rod threshold sensitivity. However, their findings did not completely justify that assertion. They say: "Nevertheless, even with red light, rod adaptation makes its appearance." It appears their illumination values are prior to filtration. The spectrum of the resulting radiance at the pupil is far from photopic. Their Maxwellian illuminator provided a fixed optical pupil size of 1 square mm.

Their caption for this figure indicated they were seeking to avoid stimulating the [XXX] color channels of the visual system. They did not specify the threshold sensitivity before the cessation of illumination. There was no knowledge of the transient associated with the P/D equation at that time. Therefore, matching a theoretical equation to their experimental data involves some unknowns. Note the large size of the test field used. The dots are their original data points. The first data point occurs at thirty seconds after cessation. This is a long enough interval for the real pupil size to have changed significantly. However, the optical pupil is controlling.

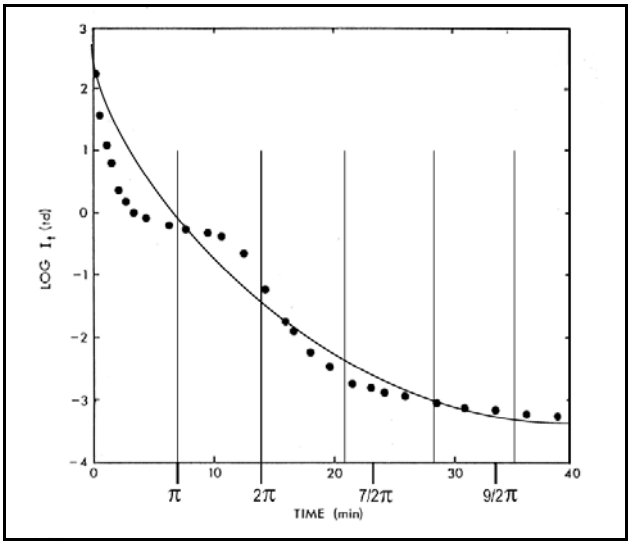


Figure 16.4.2-7 Dark adaptation curve of a human observer, following a 2 minute exposure to an adapting light of 4 x 10^5 td. The solid line represents the exponential component of the exposine function from this theory. The sinusoidal component, with a period of about 14 minutes, is well represented by the dots straddling the exponential. The dots are from Hecht, et. al. The test spot was violet, had a diameter of 3 degrees, a duration of 200 ms, and fell 30 degrees from the fovea. From Hecht, et. al. (1937)

The theoretical threshold curve defined above can be fit to this data as shown by the solid line. Note the dot at 33 minutes is above the trend line in their data and a similar situation may occur at 24 minutes. Using the values in **TABLE 16.3.5-1** and these rough approximations, the frequency of the sinusoidal component can be determined to be one cycle every 14 minutes (1.2 x 10^{-3} Hz). Looking at the time for the nulls in the sinusoidal component, accurate values of the exponential component can be obtained at 7, 14, 21, 28 and 35 minutes. With these values, it is possible to sketch in the exponential curve and estimate the height of the sinusoidal component.

Note that changes in the amplitude of the sinusoidal component effect different portions of the resultant curve differently. For a larger component, the composite curve will be lower between zero and pi but higher between pi and 2pi. This is exactly what Hecht, et. al. noticed on a daily basis in their experiments (pg. 841). This variation is indicative of the electrostenolytic reserves available at the capacitors labeled C_2 and higher.

The sketched exponential is not a true exponential because of the variation in the multiplication factor, M. This variation also distorts the sinusoidal waveform component but it does not affect the zero crossing locations. The variation expands the lower portion of the dark adaptation characteristic as the collector voltage approaches the intrinsic electrostenolytic voltage, V_{int}.

Figure 16.4.2-8 reproduces another graph from their work illustrating the effect on the dark adaptation

characteristic of location within the retina. The dashed and dash-dot lines have been added by this author. Note that a somewhat smaller, but not small, test field was used. The first data point occurs at 30 seconds from time zero and is shown as common to all of the curves. **[This curve has led to the misconception that the human eye is very insensitive in the center relative to the field. Were these curves actually normalized? Where is the center? -- What was the prior illumination level and for how long]** The curves in this presentation appear to be inconsistent, or easily misinterpreted, as drawn. Note the breakpoint, sometimes labeled the α point, between the two branches of the curve for the 10 degree field position. This break point appears to move to the right and down as the field position moves to 5 degrees. It would be expected that the same situation would continue when moving to the 2.5 degree position. However, the smooth line drawn through the 2.5 degree points is extended beyond the last obvious data point without an inflection point, although the solid point on the zero degree curve may belong to this set. The fact that the foveola, central 0.8 degrees, and the fovea, central 4.2 degrees, are structurally different from the outer retina and affect the hydraulic access to the photoreceptors in each respective region must be considered when evaluating these curves. The use of a large, 2 degree field, test area complicates this evaluation. For the 2.5 degree field location, the data for the 2 degree test area is clearly an average of a possibly rapidly changing parameter.

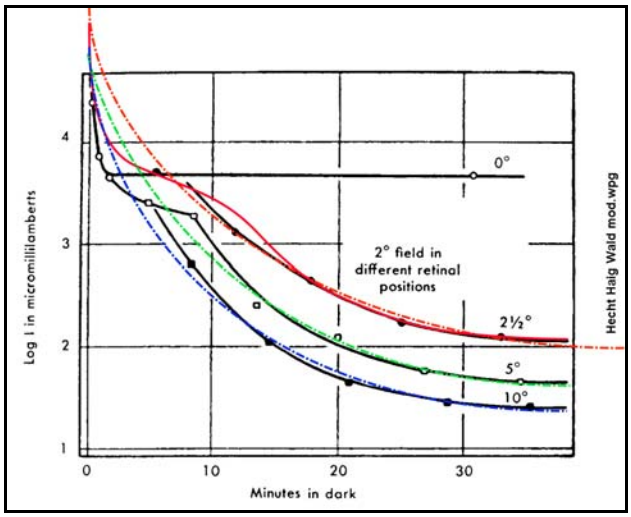


Figure 16.4.2-8 Dark adaptation at different distances from center as measured with a 2-degree field. Dashed and dash-dot construction lines added by this author. Adapted from Hecht, Haig & Wald (1935)

As an alternate interpretation of the data in this figure, consider the dashed line as an extension of the curve representing a process with a time constant of about 5 minutes. Also, consider the extension of the off-center data points back to a common time point as shown. The overall figure can now be interpreted differently based on the theory of this work. Consider the vertical scale to be linear for the moment and to represent the current through the photoreceptor distribution amplifier instead of the voltage at its pedicel. Looking at the 2.5, 5 & 10 degree lines first, the metabolic supply to the retina, averaged over a 2 degree circular field, is seen to vary slightly with distance from the center. The nominal time constant for one underlying process for all three positions is approximately 18 minutes. There is a shorter time constant that is also significant. This time constant represents a second process that affects all off-center positions in the retina. It is approximately four minutes. For retinal areas very close to the center, a third time constant is found. It is much shorter than the previous two. It is about 1minute based on the zero degree field position. In the context of this theory, it is the metabolics, delivered via the vascular system, to the Outer Segments of the photoreceptor cells of the retina that are determining these time constants. Using the analog of the vascular sub-system presented above, these three situations are addressable.

In the analog of the vascular supply network, a capacitor corresponds to the volumetric capacity of a reservoir. The charge on a capacitor corresponds to the amount of metabolic energy stored, still in chemical form except at capacitor C_A, at that location. A resistor corresponds to the impedance a metabolite encounters in attempting to flow from one reservoir to another. The voltage across either of these elements corresponds to a “field gradient” within the vascular sub-system relative to that metabolite. Using this analog, a specific time constant can be related to the different resistive and capacitive elements in the network. A given vertical distance on the graph corresponds to a change in field gradient. The maximum possible gradient would be between the reservoir represented by the blood supply and the analog equivalent of the charge on capacitor C_A. The scale on the right in the above figure illustrates this field expressed as a relative voltage. The zero point is taken at the field strength corresponding to the light adapted level and

the scale is plotted inverted. The maximum value on this scale corresponds to the fully dark adapted condition. Under this condition, no current is flowing through the adaptation amplifier, except possible thermal noise current generated within the base of the Activa. Such a current would be extremely small based on the threshold energy level of the hydronium liquid crystal. Under the fully dark adapted condition, the quiescent voltage of the Activa collector plasma would be expected to equal the intrinsic membrane potential of the dendrite and essentially no metabolic flow would be expected through the vascular circuit analog. Thus, $V_1 = V_2 = V_3 = V$.

Based on this description of the analog, it is important to discuss the ultimate value, for time --> infinity, of the curves in the original graph. Do all of the exponential curves eventually approach the cutoff field value of 4.6 or greater, or are there separate maximum fields for each capacitor in the analog? Separate maximum field values would suggest a lossy process overall or different metabolic constituents serving different locations in the retina relative to the center. Both of these conditions are unlikely. It is more likely that all of the exponential curves eventually approach the maximum value, taken here as between 4.6 and 5.0. Under this assumption, the zero degree line in the figure is not horizontal. It is similar to the dashed line in that it exhibits two time constants, the first one remains near one minute but the second is measured in hours, possibly many hours.

The following scenario is merely suggestive pending further laboratory experimentation. It omits any time constants, such as those associated with the P/D equation and the operation of the iris, that are shorter than 30 seconds. The assumption is also made that only one unit of energy is extracted from a single chemical reaction within the putative glutamate metabolic cycle in support of the electrostenolytic process. This may be conservative by a factor of two or three but simplifies the discussion. This scenario will also be based on the assumption that the photoreceptor cells are provided electrical energy via the choroid artery and the RPE. An alternate assumption would be that they are supplied via the artery supporting the signal manipulation part of the retina and involves diffusion through the Outer Limiting Membrane. Little is known about the exact nature of the IPM and the RPE in terms of their precise dimensions and permeabilities to the nutrients of interest. It is logical to assume the shortest time constant corresponds to the smallest physical size elements in the analog. This would correspond to the capacitor, C_A, combined with the impedance of the electrostenolytic process on the wall of the cell. This time constant corresponds to the common part of the graph in the above figure, and will be taken as one minute for all of the photoreceptors regardless of their position in the retina. The field intensity associated with this time constant would be about one unit on the right hand scale and corresponds to the field difference between V and V_1. The second time constant appears to be closely related to the position of the test field relative to the center. The field strength between V2 and V1 of the vascular supply is quite small at the zero degree position, rising to about one equivalent volt across the capacitor at V_1, corresponding to the 2.5 degree position (the dashed line). This would correspond to a relatively long physical path for the metabolites. The length of the path would be longest at the center of the retina. This would suggest a structure for the diffusion path similar to that found within the signal manipulation layers of the retina where there is little or no vascular flow to the macula in the vicinity of the fixation point. The region near the fixation point is served only by bulk diffusion. Near the edge of the fovea, the time constant has been reduced to between 20 and 30 minutes, primarily due to the shorter diffusion path before interfacing with the conductive components of the vascular sub-system. Once beyond the fovea, the metabolic supply to the photoreceptors takes on a different character. The local diffusion path becomes less and less important, both in capacity and permeability. The primary performance determinants become the conductive flow of metabolites to the vicinity of the photoreceptor, probably defined by the porosity of the conduit walls and the capacity of the vascular bed adjacent to the vascular supply. The time constant of the 2.5, 5 & 10 degree field positions appears to be quite similar. However, this is probably misleading since the curves are drawn to be horizontal at the 40 minute point. Subject to second order effects, such as the variable thickness of the vascular conduit walls with position in the retina, it is likely that these three curves eventually converge at a relative field strength of between 4.6 and 5. This would suggest that the 10 degree field position has the shortest time constant and the time constant becomes longer until it equals or exceeds the time constant of the dashed line.

Note that in a circular spatial field, the percentage of the area within a given radius taken up by a circular test spot becomes larger. This relative size appears to impact the observed performance of the retina. This suggests that the dark adaptation performance of the retina is a significant function of the size of the test spot used. A considerably different characteristic might be obtained using a simulated point image instead of a finite diameter test spot.

Considerably pathological data has been reported relative to the dark adaptation characteristic. This data will be referenced in Chapter 18. The data appears to be consistent with the above scenario.

16.4.2.2.3 The nominal characteristics for light adaptation

The light adaptation characteristic has not been a subject of great laboratory interest because it occurs so rapidly and is hard to measure. Even with a complete schematic of the adaptation amplifier and its power supply, the theoretical description of this transient is complicated. Light adaptation involves the same circuit diagram as shown in [**Figure 16.3.5-6** XXX]

For a large transient input signal, the overall light adaptation characteristic exhibits three time constants. The first is the rise time of the P/D equation. The second is the time constant of the Activa collector circuit. The third time constant is that of the iris.

16.4.2.2.4 Evaluating the parameters of light adaptation

Poot, Snippen & van Hateren have provided unique psychophysical data on the light adaptation characteristic of the human eye at a wavelength of 563 +2,-0 nm. (FWHM =30 nm.)[56] and a current list of citation. Several different test scenarios were used. However, their work did not address all of the time constants and processes discussed above. In particular, several of their figures show values for negative circuit delays under some conditions because they did not compute the time delay as a function of illumination in the P/D equation. They did focus on relatively small illumination increments representative of a real life visual encounter. However, they did not correlate their absolute illumination levels to the operating levels of the iris. Since the experiment times were short, any change in the pupil diameter would be small. However, it could contribute to the slope of the data points in the 200 to 800 ms. interval of their data and the ultimate threshold values on the right in several of their drawings. Their procedure defines a very short initial adaptation time of only 30 seconds. Were the subjects able to look out a window, or at any significantly different light source before or between test runs? Neither did their work discuss the longer time constant of the iris. Their statement: “Light adaptation is nearly complete within 100 ms after luminance decrements” only applies to small decrements. The use of a 10 ms test signal for determining the perceptual threshold of the subjects is quite short relative to the transient response of the visual system of the same subject to this stimulus. Based on this work, the terminology in paragraph three of their introduction can be expanded and more precise conclusions drawn. The title of their paper succinctly describes the performance of the combined transduction and adaptation circuit, over the total interval and at the incremental interval they examined. However, the situation is more complex than they present.

Earlier discussions have suggested that the perceptual threshold of the human eye to short pulses in one chromatic photodetection channel could be described in terms of a fixed differential voltage threshold at the pedicel of the photoreceptor cell. This threshold can be described in terms of a differential current through the distribution amplifier and the impedance of the axolemma of that cell. That impedance is represented by a diode. Under less than saturating conditions, this differential current is equal to the negative of the differential current through the Adaptation amplifier of the same cell. If these propositions hold, the perceptual threshold can be determined by computing the differential signal amplitude, a differential current, at the emitter of the adaptation amplifier. For the conditions investigated by Poot, et. al., the amplitude of, and time delay related to, this differential current is a transient function of the input signal level, the state of the chromophores in the P/D process, and the state of the adaptation amplifier. A question that must be addressed is what is the precise composition of the transient required to reach the threshold level. The best data available suggests that the time constant of the adaptation amplifier is about 300 ms for decrements in illumination and considerably shorter than 300 ms for increments. Based on the P/D equation developed in this work, the time constants for an increment in illumination varies greatly with illumination level. However, the time constant for decrements is more constant. Based on data available for dark adapted conditions, the decay time constant of the P/D process is about 1.5 seconds. The rise time is given by $\sigma F \tau$ (a different τ than in Poot) and is generally much less than 1.5 seconds. The P/D equation also exhibits a time delay that is related primarily to its state of illumination before any change in that illumination.

---

[56]Poot, L. Snippe, H. & van Hateren, J. (1997) Dynamics of adaptation at high luminances: adaptation is faster after luminance decrements than after luminance increments.

It appears that a more detailed calculation of the sum of these transients, which is beyond the scope of this work, can provide a theoretical basis for the data of Poot, et. al. Comparing theory and data would be easier if data were available for only one chromatic channel.

Figure 16.4.2-9 shows a figure from Poot et al. and another from Uttal (page 502). In both cases, the on-transient is present but difficult to evaluate because of the time scale. Neither time scale extends long enough to evaluate the decay time constant precisely. [xxx expand discussion]

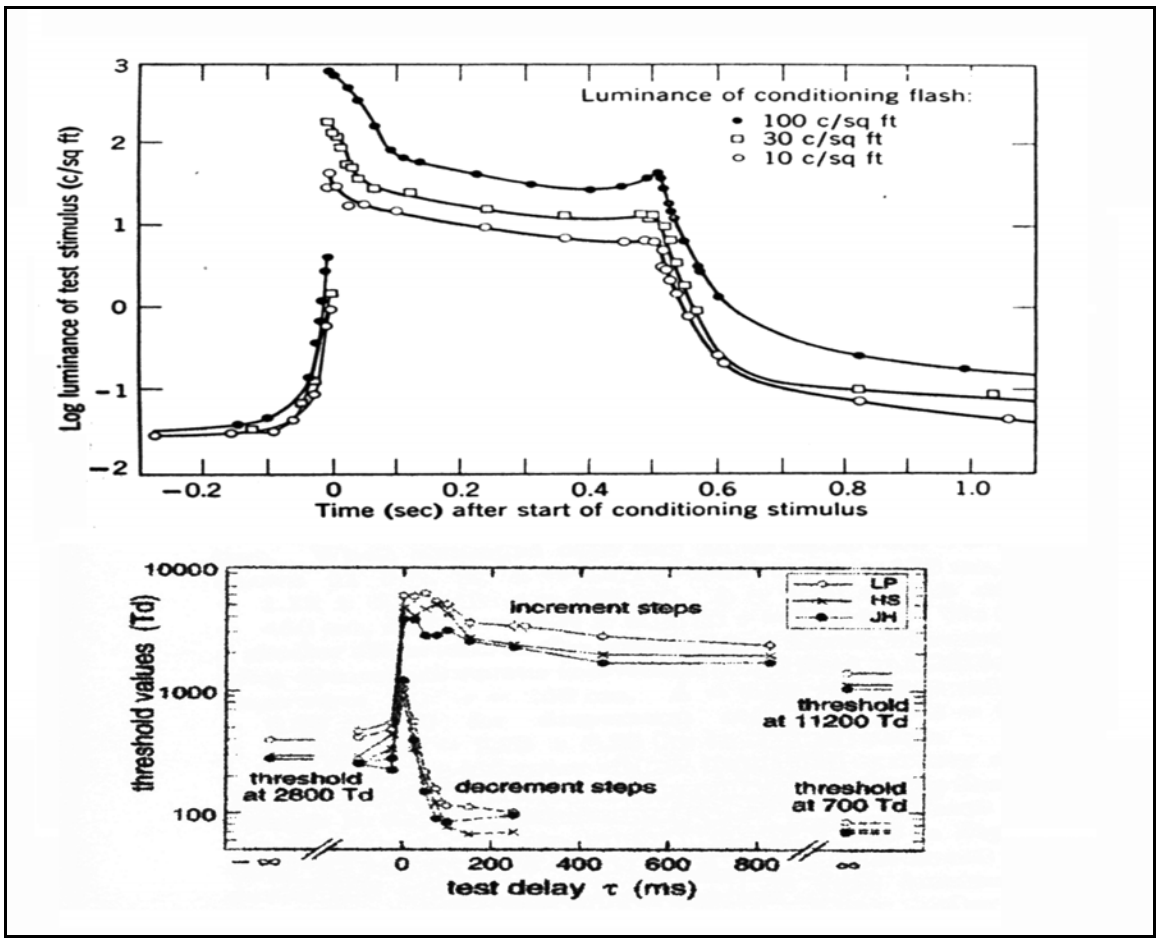


Figure 16.4.2-9 REWORK Theoretical light adaptation characteristic. Compare to Poot, et. al. (bottom) and Mueller in Uttal fg 7.15 (top). The time resolution and extent in Uttal is not appropriate for determining the initial rise time or the actualy decay characteristic. Similar comment applies to the Poot et al. frame.

16.4.3 The generator potential at the PC pedicle EMPTY

16.4.4 Temporal equation of signal manipulation in stage 2 EMPTY

16.4.5 Temporal equation of signal projection in stage 3

The signal projection system, used primarily, if not exclusively in *Chordata*, is designed to support the maximum possible agility of the eye relative to the body. The primary requirement is to reduce the stiffness of the optic nerve by reducing the number of individual neural circuits within it. This is achieved by spatial encoding and decoding (analog) in conjunction with pulse encoding and decoding (action potentials). The algorithm used in the pulse encoding/decoding function is frequently described as a codec. The fundamental circuit used is illustrated in **Figure 16.4.5-1**. The figure represents an expansion of the luminance and chrominance paths shown in the Fundamental Circuit Diagram of **Section XXX**. The upper path represents a luminance channel in the projection stage. B and C represent a pair of chrominance channels. B or C could be used equally well to describe a polarization, N --channel (**Section 16.2.4.3**). Note the use of a different ground symbol in the return leads of the Nodes of Ranvier, and the 3rd synapses on more tenuous grounds, consistent with their different electrolytic location in the visual system from either the retina or the brain proper.

All of the signals arriving at the input terminals on the left are of Class E. They represent signals from the signal manipulation stage of the retina. Only a single 2nd generic synapse is shown at the entrance to each channel. However, it represents hundreds to thousands of synapses associated with the input to each ganglion cell. These synapses receive signals from the myriad fibers of the inner fiber layer. The mapping of the signals from the second stage neurons to the ganglion cells of the third stage represent the spatial encoding of the retina. At the output of the projection system, a similar mapping of signals from the individual signal channels to the 3rd generic synapses connecting to the initial stellate cells is found. In the projection system, these two mapping functions are paired and together provide both the spatial encoding and decoding function. In the overall architecture of the visual system as known today, the combined mapping and unmapping are transparent.

A distinction must be clearly made between the encoding performed in the signal manipulation stage compared to the signal encoding performed in the inner fiber layer (portion of the inner plexiform layer closest to the ganglion cells) leading to the signal projection stage. The former is part of the signal processing related to cognition and perception. The latter is primarily to provide ocular agility. This

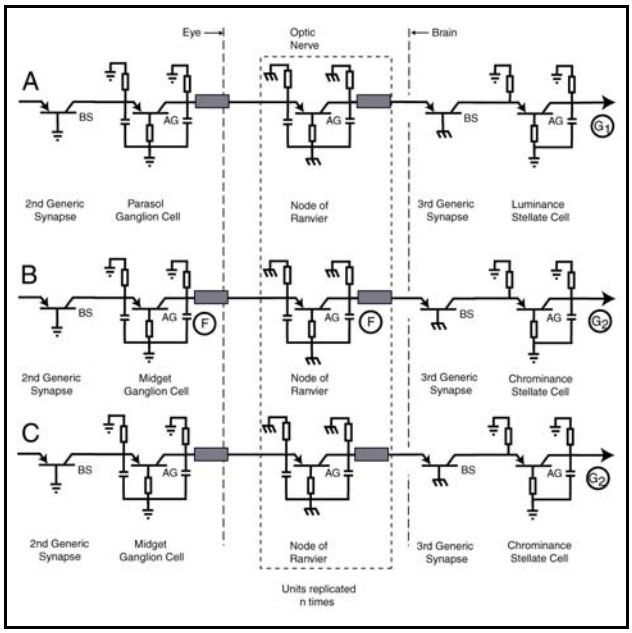


Figure 16.4.5-1 Fundamental circuit of the signal projection sub-system.

differentiation has a bearing on the comments of Mullen & Kingdom in Gouras[57] working in the psychophysiological arena. There text defines both a luminance (cone additive) and a colour (cone opponent) signaling channel supported by "one mechanism." They then make the broad statement that "The scheme...is only useful if the signals can be subsequently decoded (demultiplexed)." This work suggests that statement is too broad and not adequately focused . Only the signal projection portion (primarily spatial with regard to retinal position) of the encoding needs to be subsequently decoded. The signal manipulation encoding (or differencing) is an intimate part of the on-going cognitive process. As an example, *each of the two* chrominance channels provide, in encoded form, a linear discrimination function as a function of spectral wavelength. Nothing is gained by decoding this signal back to the photoreceptor level. Similarly the encoded luminance channel represents the perceived luminosity function, no decoding of this signal is needed or appropriate.

Each of the signal paths has a ganglion cell associated with it. However, the parasol cells are biased differently than the midget cells. The parasol cells are biased to cutoff. In the absence of a monophase signal at their input, these cells do not generate an action potential. The midget cells are biased to produce a continuous series of action potentials in the absence of any biphase input signal. While only the waveform at the output of the ganglion cell in channel B is labeled, the output of all of the ganglion cells are represented by a class F waveform. However, the frequency between these pulses is a separate parameter in these channels. The shaded boxes represent the transmission lines associated with each ganglion axon and interneuron segment. As the class F waveforms propagate along these transmission lines, they are degraded in amplitude and phase. Before they become useless, each signal is applied to the input of a Node of Ranvier. The Activa within the node is usually biased to cutoff. It does not generate any action potentials without excitation. This node regenerates the action potential, not by amplification but by regeneration. Technically, the new class F waveform now exhibits the characteristics associated with the circuit parameters of its node. Since all of these nodes are nearly identical, all of the output waveforms appear to be identical action potentials except under pathological or experimental conditions. The distance that a signal can travel along the transmission line of an axon or interneuron is typically less than 2 millimeters. The dotted box indicates that the Nodes of Ranvier are replicated n times in order to transmit the signals from the rear of the retina to the input structures of the brain. The resulting value of n may be 25 or more in the visual system (and much higher in other systems extending over the length of the animals body). The value of this number illustrates the importance of regeneration, rather than amplification at the Node of Ranvier. The total time delay associated with each of the n replications is given by the delay of the signal in the delay line of the circuit and the delay introduced by the regeneration process. The latter delay is the dominant delay and leads to the slow signal velocity often quoted in the literature (erroneously) as the velocity of the neural signal.

> There are a few references in the literature to the bifurcation of an axon. Based on this work, it appears that there is a need for more precise terminology in this area. To bifurcate an axon between Nodes of Ranvier or synapses would suggest the current in the axoplasm is free to divide between two or more signal paths in a manner controlled by the terminal impedances of the branches of the axon. This leads to an apparently undesirable situation. It appears more likely that any bifurcation of an axon actually occurs at a voltage node. At this point, multiple synapses can connect to multiple orthodromic neurons without complication. Alternately, a Node of Ranvier with one input structure and two (or more) output structures can be formed. Under this interpretation, a single axon does not bifurcate, except at a Node of Ranvier. It is the Node of an axon that bifurcates.

Note that all of the Activas in this figure are described as Type AG. These devices are all similar and exhibit essentially the same current carrying capacity. It is the biasing arrangement provided by the associated components that determine the operation of the circuit. This is particularly obvious in the case of the stellate cells where no capacitance is shown in the emitter circuit of these neurons. These circuits are not designed to generate, or regenerate, action potentials. They are designed to decode action potentials. They will only oscillate due to the addition of capacitance to the circuit through laboratory intervention.

The circuitry of the brain is beyond the scope of this work. However, it is useful to discuss several points. The simplified circuit believed to be associated with the first signal processing stage of the brain are shown on the right for each channel. These circuits appear to be those of the stellate cells found at these locations.

---

[57]Mullen, K. & Kingdom, F. (1991) Colour contrast in form perception. *In*, The Perception of Colour, Gouras, P. *ed.* Boca Raton, FL: CRC Press, pg. 206

These circuits have the initial task of recovering the analog waveforms represented by the encoded action potentials. In this sense, they are paired with their corresponding ganglion cell. The ganglion cell encodes and the stellate cell decodes, hopefully in an overall transparent manner. However, this goal is not achieved as will be discussed below. A more complex conceptual decoding circuit was presented in **Section 11.7.2.1**. There are a variety of series and parallel implementations of decoding circuits that could be used in the architecture of the brain. There are three primary considerations, circuit economy, circuit performance and circuit delay. **[Figure 11.7.2-2]** showed a parallel implementation that minimizes delay in the "alarm" signal output. In the outputs to the cognition circuits, quality of reproduction may be more important than time delay.

The output signal of the luminance channel is labeled as Class G_1. This subclass is characterized by an output level of "zero" volts in the absence of an input signal to the luminance channel, e. g., no excitation of any of the chromatic photoreceptors associated with that channel. By contrast, the two luminance channels exhibit a nominal output voltage in the absence of any information in the chrominance channels. They are labeled Class G_2. This output voltage may increase or decrease depending on the signals created in the signal manipulation stage.

It is the capability of the net signal amplitude in the chrominance channels to increase or decrease in response to an input photon flux that leads to the assumption that there are OFF type signals in the luminance channels of vision. Whereas, a signal can be located, by probing, that goes in the OFF direction when the illumination is increased, that signal is not from a luminance channel. It is from a chrominance channel. This can be demonstrated in a number of ways. First, the maximum action potential frequency in such a channel will not approach the maximum frequency in a similar luminance channel. The maximum will normally be that of the quiescent chrominance channel for that species, typically 30 Hz. Secondly, the maximum frequency of the signal at that probe location can be increased by introducing additional illumination of a different spectral content, thereby demonstrating the subtractive nature of the chrominance channels.

Figure 16.4.5-2 describes the neurological morphology of the input connections to the brain[58]. The primary decoding neurons in the brain are found among the stellate cells, with granule cells a sub-category of stellate cells. It appears that the stellate cells may act as either decoders or as regenerators with inputs from several individual efferent nerve fibers. The pyramid cells appear to act as re-encoders, sending signals to distant parts of the brain via "association fibers." The individual layers are subdivided and more path names are given in Hubel[59]. However, less detail is provided than in Noback.

[58]Noback, C. (1967) The human nervous system. NY: McGraw-Hill, pg. 233

[59]Hubel, D. (1988) Eye, Brain, and Vision. NY: W. H. Freeman pg 99-104

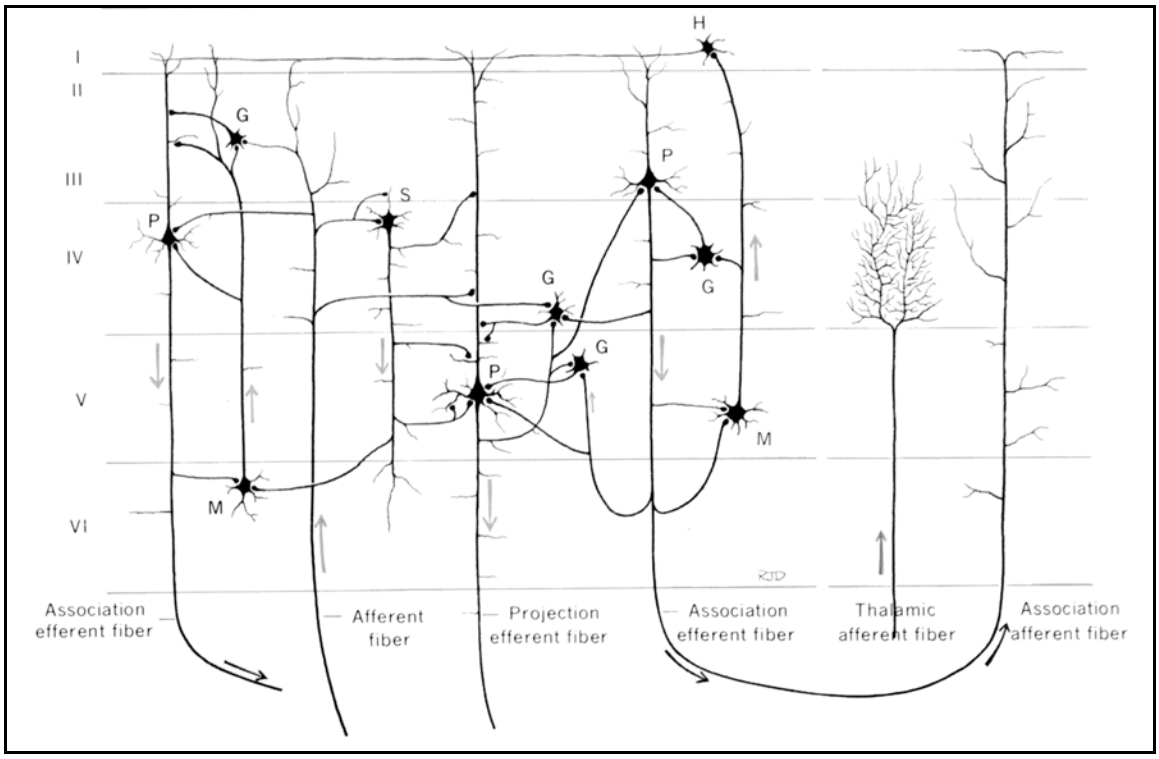


Figure 16.4.5-2 CR Topology of the brain's input and transfer circuitry. EXTEND CAPTION The top of the figure represents the exterior surface of the cerebral cortex. From Noback (1967)

The codec circuits associated with signal projection play a minor but perceptually significant role in vision. The role is more significant in human vision for a subject living in an advanced industrial society. The codec employed in the luminance channels is well behaved and can be considered to have a transfer function of 1.0. There is some data in the literature showing the ganglion cell employs a threshold below which no action potentials are generated. It appears that this is technically true but the threshold level is very close to the noise level in the channel and does not affect the functioning of the luminance channel under operational illumination conditions. The codec used in the chrominance channels, and by analogy, other signal differencing channels is symmetrical with respect to input signal amplitude. However, with large amplitudes, the time interval between action potential pulses can become equal or longer than the time interval between changes in elements of the input scene. This can result in significant errors in the chromatic content of the scene perceived by the individual. The result is a long list of special chromatic oddities reported in the literature. The basic process is illustrated in **Figure 16.4.5-3** and will be discussed in more detail in **Section 17.6.3.** The effect of this asymmetry is noticed particularly in the slower perception of S- and L-channel signals.

This work continues to find that the distribution of chromatic photoreceptors in the retina and the interconnections within the signal manipulation and signal projection stages are not randomly arranged but are more highly optimized through genetic or other control mechanisms.

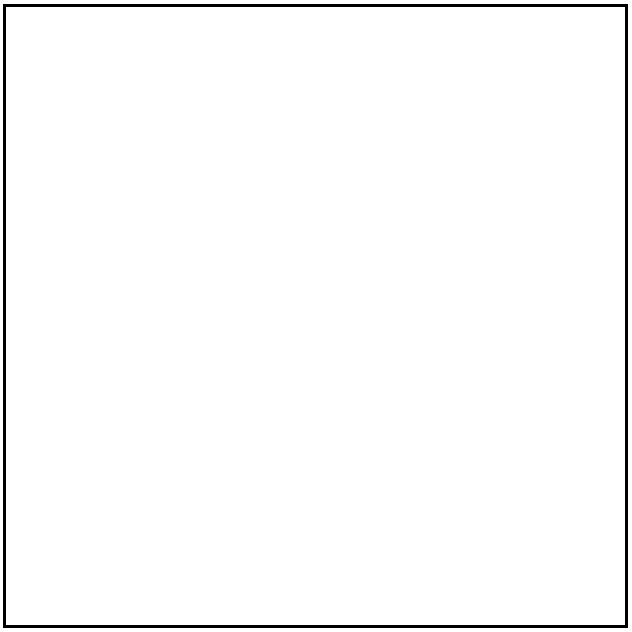


Figure 16.4.5-3 Asymmetrical operation of the signal projection codec in the chrominance channels of vision

16.4.5.1.1 Nominal waveforms of signal manipulation

Because of the variety of waveforms found in the signal projection circuits of vision, it is useful to review them in order to understand the frequency response and time delay of those circuits. **Figure 16.4.5-4** illustrates a selection of these waveforms in conceptual form. It is difficult to be specific in this area because of the large number of variables involved. (A) and (B) of the figure relate to the ganglion cell of the retina. (A) shows the fundamental circuit of the ganglion cell, including the delay line between the collector terminal of the Activa and the input terminal of the first Node of Ranvier. (B) shows the pulse to pulse time delay of this circuit as a function of the voltage between the emitter and the base of the Activa. If the ganglion cell us used in a luminance channel (therefore defined here as of the parasol type), the device is usually biased at cutoff (location indicated by a p on the time scale). Under this condition, the device does not generate any action potentials until a signal is applied to the input terminal. It then generates one action potential, waits a short interval, known as the reactive period, while it re-establishes its quiescent conditions and then generates a second action potential. The time between these pulses, and subsequent pulse pairs, is determined by the emitter to base voltage following the reactive period. This period can be as short as a nominal 10 ms. or indefinitely long. Under fully dark adapted conditions, it is possible that no pulse will ever be generated by a luminance ganglion cell. Alternately, a few noise pulses may occasionally cause an action potential to be generated. Further laboratory analysis is needed to determine the actual situation in a particular species.

If the above cell is used in a chrominance (or other differencing) channel, the device is usually biased to generate a continuous series of action potentials with a period of 33 ms. in the absence of any input signal. This bias is shown by the letter m corresponding to the label of midget ganglion cell. When so biased, the time interval between individual pairs of action potentials is determined by the instantaneous voltage, V_{EB}, following a brief reactive period. Since this voltage can be biphasic, the interval between pulses can decrease or increase as shown on the graph. In both cases, the instantaneous frequency of the pulse train can be computed as the reciprocal of the interpulse time interval. However, remember that a single pulse is significant in this system.

> It is important to note that it is the time delay between the individual action potentials that is the fundamental parameter in the encoding of the signal information. The time of occurrence of the first (or only) action potential in a sequence is of critical importance to the alarm function. The spacing between the subsequent pulses in the sequence carry critical information for the cognition functions of the brain. However, it is common to talk in terms of the instantaneous frequency of the pulses. To calculate or speak in terms of the average frequency is to lose sight of the process involved. No frequency can be associated with this single pulse.

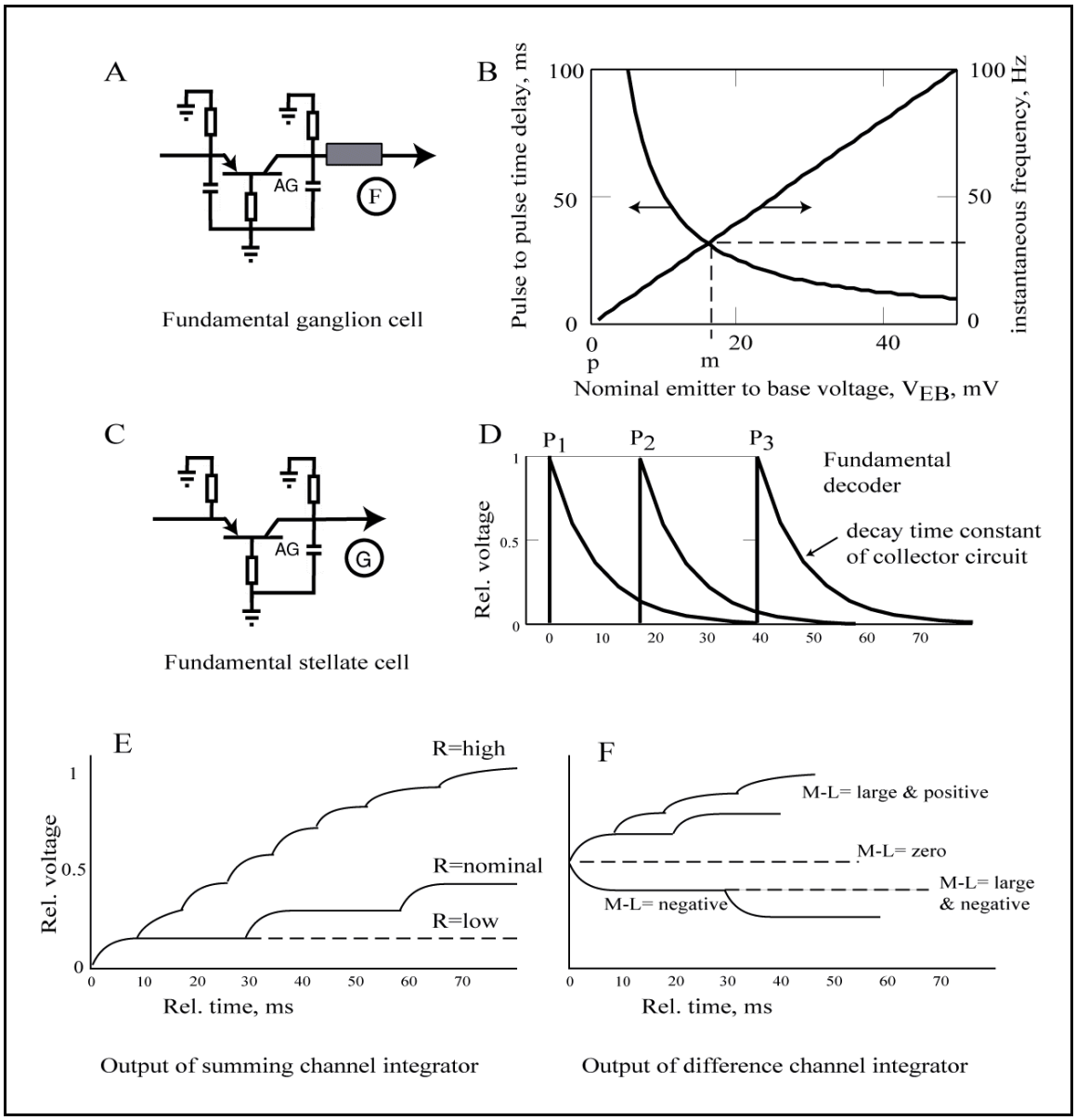


Figure 16.4.5-4 Waveforms associated with the signal projection stage. See text for details.

The pulses generated under either of the above scenarios are transmitted down the axon of the cell to the first Node of Ranvier. The time for the signal to travel to this location is controlled by the phase velocity of the axon when considered as a transmission line. This subject is discussed in the next Section.

At the Node of Ranvier, the Activa is normally biased off, similar to the parasol ganglion cell, regardless of whether it is in a luminance or differencing channel. It does not generate any action potentials on its own volition. It acts strictly as a re-generator. However, the action potentials it generates looks virtually identical to the action potentials created by the previous Activa. As a result, a virtually identical series of pulses are created delayed in time by the phase delay of the previous axon and the regeneration delay of the Node of Ranvier. This process is repeated at each Node of Ranvier until the 3rd generic synapse is reached at the entry point to the cortex. Each Node of Ranvier and its orthodromic interneuron introduces a delay that is dominated by the regeneration delay. This regeneration delay is very similar to the generation delay associated with the first Activa in the ganglion cell. Thus, the total delay associated with a particular

ganglion cell is approximately N+1 times the regeneration delay of a single Activa circuit. It is only necessary to calculate the phase delay of the axon and each associated interneuron when very long neurons are involved.

In summary, the time interval between the action potentials of the ganglion cells depends on how they are biased. In the case of the parasol cells and the Nodes of Ranvier, this interval can be infinite under dark adapted conditions. The time interval in the case of the midget ganglion cells is nominally 33 ms. under similar conditions.

(C) and (D) of the figure illustrate the initial decoding process associated with the signal projection stage. After the 3rd generic synapse, the signal is applied to a stellate cell (as defined here). The initial circuit of this cell is very similar to that of a ganglion cell, except its combination of circuit elements and bias conditions does not lead to oscillation or re-generation. The Activa in a stellate cell associated with the luminance channel is usually biased at the cutoff point like the Activa in a parasol cell. The capacitance in the collector circuit acts as an integrator. Upon the arrival of each action potential at the stellate cell, a pulse of current is passed through the Activa, thereby reducing the voltage on the capacitor during the duration of the action potential. Following that interval, the circuit begins to return to its quiescent state. This is indicated by the exponential decay characteristic associated with the first waveform, labeled P_1 in (D). What happens subsequently is not important in the alarm channel of the visual cortex. However, the arrival of subsequent pulses is important to the cognitive portions of the visual cortex. The cognitive channels need to know the amplitude of the original illumination. This information is encoded in the pulse intervals of the pulse train. As indicated earlier, there are a variety of circuit configurations that can extract this information. In this example, let the stellate cell be followed by a second cell also arranged as an integrator with a longer time constant. This circuit will integrate the charge from the previous circuit as represented by the area under the pulse waveforms of (D). The result is shown conceptually in (E). Three waveforms are shown, one for a high level of signal in the luminance, R, channel, one for a nominal signal and one for a low signal. Notice that all of these waveforms are identical following the first pulse. They then increase in amplitude in a step like manner upon the arrival of additional pulses. Each step is exponential in form. For very low level signals, the time required can be quite long. It may be on the order of seconds. (E) illustrates the difficulty of defining a specific time constant for each of these waveforms.

The operation of the stellate cell for a differencing channel is shown in (F). Here, the cell is still biased to cutoff. However, it received a continual stream of action potentials. The result is a quiescent voltage other than zero on the capacitor in the collector circuit. The voltage on the capacitor can be increased by causing more pulses to arrive at the emitter per unit time. It will decrease if fewer pulses arrive per unit time. This circuit exhibits several unique features. Whereas a higher signal amplitude at the input of the ganglion cell in a luminance channel always generates more frequent action pulses, this is not true in a differencing channel such as the chrominance channel. In that type of channel, an increase in the M-channel signal causes an increase in instantaneous pulse frequency and an ultimate increase in the output voltage of the integrator following the stellate cell. The greater the M-channel illumination, the higher the voltage and the shorter the time to reach a steady state value. This is not the situation in either an S- or L-channel connected to the inverting terminal of a neuron in the signal manipulation stage. For that channel, an increase in illumination suppresses the generation of action potentials. With fewer action potentials arriving at the stellate cell, fewer increments of charge are passed to the integrator and its voltage decreases. The higher the illumination in this channel, the more time between the action pulses. The result is shown in the lower part of (F). For a nominally negative signal at the input to the ganglion cell, the output signal decreases in steps. The steps occur less often than for positive values of the signal, M-L (or M-S). For a maximally negative signal, the generation of action potentials is nearly completely suppressed. There is no threshold circuit in the ganglion cells of the differencing channels. As a result, there is no initial pulse representing the start of a high level of L- channel illumination. The pulse train actually terminates. No pulse will be generated by the ganglion cell, and received by the stellate cell, for a considerable length of time. Thus, no change in the signal level of the chrominance channel will result. In this situation, the visual system may perceive a luminance signal but receive no new chromatic information to associate with it within a significant interval of time. This situation is a gold mine to a magician.

16.4.5.1.2 The axon as a transmission line

The overall transmission line used in signal projection is more complex than normally indicated in the literature. It consists of a long “axon” containing a series of pulse regeneration circuits separated by

individual sections of high quality passive transmission line. An unusual feature, compared to similar transmission lines used by man, is the fact that the delay associated with each regenerator is greater than the delay associated with the adjacent section of passive transmission line. In fact, the regenerator delay dominates the calculation of the group velocity associated with the signal information.

16.4.5.1.3 The ganglion cell circuit- - (time div. modulation in two forms) EMPTY

16.4.5.1.4 Signaling parameters of the projection stage EMPTY

16.4.6 Oculomotor equations of vision

These equations of the oculomotor system are presented in detail in **Section 7.3**. The performance of the iris is presented in **Section 2.4.3.1**. It is stressed in that section that **the role of the iris is not to control the intensity of radiation falling on the retina**. The primary role of the iris is to optimize the performance of the physiological optics to the available light level. It does this by changing the F/# of the optical system and thereby matching the spatial resolution of the visual system to the available signal to noise ratio.

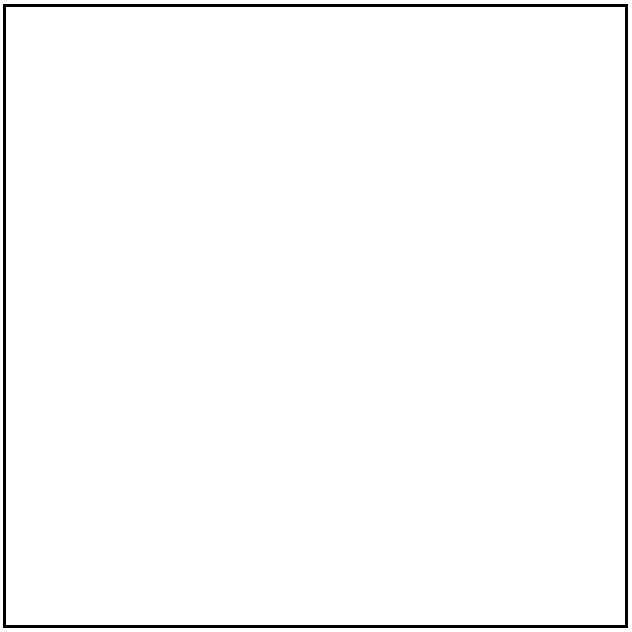


Figure 16.4.5-5 Selected waveforms of the signal projection stage. XXX XXX

16.4.7 Important postulates of this theory

Several postulates appeared during the development of the material in the following sections of this chapter. Presentation of these postulates here will aid in the understanding of that material and may play an important part in further research.

16.4.7.1 Location of the plane of perception

As shown in the following sections, the performance of the human eye, as determined by end to end psychophysical experiments, can be very accurately and precisely predicted mathematically using the equations of vision evaluated at the input to the ganglion cells. An exception being the flicker phenomena resulting from the asymmetry of the midget ganglion cells. This situation is intellectually satisfying and important experimentally if it is true that the encoding, transmission and decoding functions as a group are in fact designed to appear transparent to the signal information.

Based on the premise that a perceptual plane can be defined immediately prior to the ganglion cells for most purposes, it is possible to delineate further the functional division of labor within the visual system. The conclusion can be drawn that essentially all signal processing occurs in the retina. The remaining functions of full frame memory manipulation, cognition, and response are activities performed within the cortex. This conclusion has the importance of a postulate in this work.

A corollary to this postulate is that the transparency of the encoding, transmission and decoding functions as a group is not perfect. The group introduces an asymmetry in the chrominance channels, and probably any geometric channels, that must be accounted for separately.

16.4.7.2 Equality of signal amplitudes at pedicles

The gain coefficients, K, in the above equations are quite complex. They can be separated into a number of individual components. The primary separation is into two components, the constant amplitude portion and the variable portion. It is then possible to separate the fixed amplitude portion into two components; the fixed component of each photodetection channel determining the magnitude of the signal at each voltage node, and the fixed component relating the various voltage nodes to the inputs to the Activa of the individual processing amplifiers. Symmetry would suggest that the nominal fixed gain of each photodetection channel would be the same. Furthermore, the success of predicting the wavelength discrimination function of the human eye using equal photon fluxes per unit wavelength applied to each channel would suggest that the coefficients controlling the signal amplitude at the voltage nodes were the same.

It is therefore postulated, subject to further experimental confirmation, that the nominal fixed gain coefficient for all photoexcitation channels are the same. Any variation from this feature is associated with the variable gain characteristics specifically related to that individual channel. This postulate is also intellectually satisfying in that it focuses attention on the unique characteristics of the L-channel. Under steady state conditions, ranging over many orders of magnitude, it appears that the fixed coefficients associated with the UV-, S-, and M- channels remain equal and the variable coefficients associated with adaptation track each other remarkably well. It is only the L-channel that fails to track the UV-, S- and M-channels. This failure is directly relatable to the square-law process associated with the conversion, on a quantum basis, of excitons to free charge carriers.

A corollary to the above postulate is that the fixed gain coefficients associated with the input circuits of all chrominance signal differencing circuits are the same. The Fixed gain coefficients associated with the input circuits to the luminance signal summing circuit may be different in order to generate the composite luminance function.

16.4.8 Extensions to the fundamental equations

The above equations do not adequately describe the real performance of the visual system. They need to be extended to include the effects of the physiological optics and of the variable gain associated with the adaptation amplifiers. Introduction of the variable gain of the adaptation amplifiers leads directly to the color and brightness constancy of the visual system.

16.4.8.1 The variation in performance due to the physiological optics

While the above equations correctly describe the performance of the visual system beginning with the photoreceptor cells, they do not include the absorption of the neural layers of the retina (the macula) or the absorption due to the physiological optics prior to the retina.

16.4.8.2 The variation in performance due to the adaptation amplifiers

The variation in the gain of the adaptation amplifiers and their specific circuit configuration leads to a gain term that is inversely proportional to the incident irradiance over their operating range. This range is defined as their photopic range. When all of the adaptation amplifiers are operating within their photopic range, the complete visual system is operating within its photopic range.

16.5 A summary of signaling in the visual system

It is now possible to assemble sets of waveforms that can describe the temporal aspects of the signaling channels in animal vision. Each set provides a comprehensive view of the gross changes in the amplitude response and the time delay associated with the signal as it travels along the neural network. The notation developed earlier will be convenient. **Figure 16.5.1-1** shows a simplified top level schematic with the additional notation.

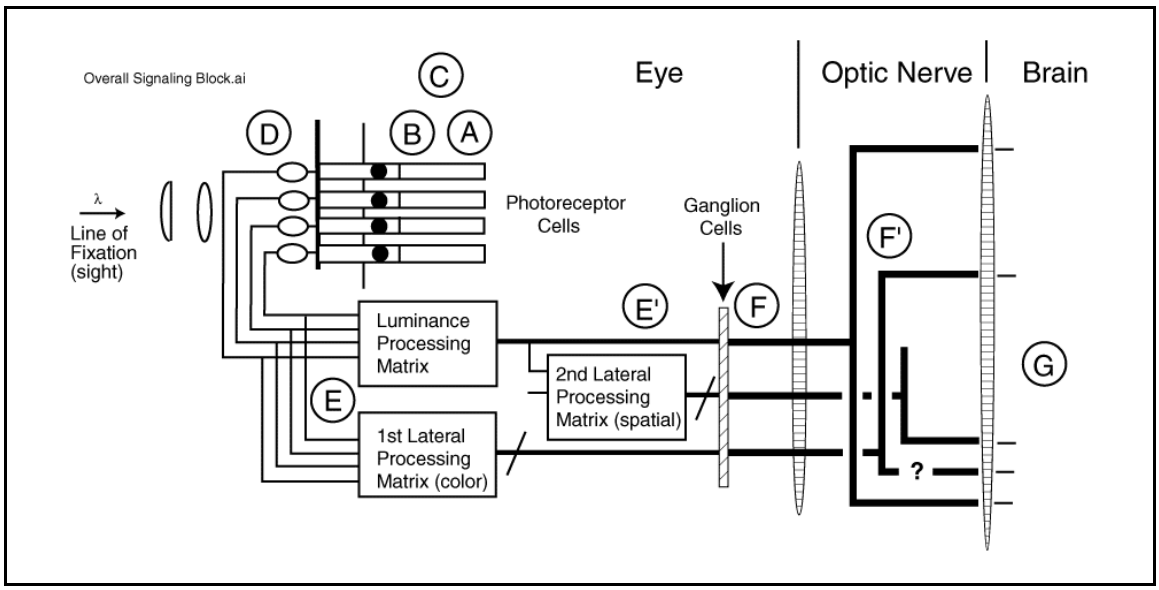


Figure 16.5.1-1 Top level schematic of the visual system of Chordata and the waveform Class associated with each location.

The signal created by the de-excitation of excitons at the disks/dendrites interface within the OS is not measurable with conventional electrical probing techniques. The electrical signal created in the base terminal of the adaptation amplifier is also probably not measurable with the current probe technology. Although this signal is a current of less than 1000 Hertz bandwidth, it is extremely small and consists of both an electron and a hole current. The net current is typically less than a micro-pico-ampere (10^{-12} amperes). The net signal might be theoretically measurable using Hall Effect techniques, however, it would be extremely awkward to do this for a single cell. At this location, the signal is described by the complete P/D equation. The complete equation includes both the amplitude response and the phase response, which describes the time delay of the preceding circuit. The waveform of the signal at this location will be called a Class A waveform.

There are two waveforms measurable at the emitter and *in the region of the* collector of the adaptation amplifier that are similar but distinctly different. The Class B waveform is found at the emitter terminal of the adaptation amplifier, which is shared with the emitter terminal of the distribution amplifier of the photoreceptor cell. This electrical waveform is essentially an amplified copy of the Class A waveform created by the photodetection process. At the detailed level, there may be a slight variation due to the difference in lifetimes between the electrons and holes in the base region. The waveform is unidirectional as discussed below.

The waveform *in the region of the* collector is differentiated from the waveform at the collector terminal, the axoplasm of the adaptation amplifier because of accessibility. Recall that the adaptation amplifier consists of a group of distributed base Activas formed by the microtubules within the fissures along the sides of the disk stack. Determining how to capture the collector current emanating from the microtubules separately from the electrostenolytic current supplying both the collector and the capacitance of the membrane wall of the individual microtubules will be left as an exercise for the student. Until a method of separating these two currents appears, the waveform associated with this combined current from all of the microtubules of a given disk stack (or OS) will be labeled Class C. It is the waveform collected by the suction micropipette technique *before* the electrostenolytic supplies are exhausted and the circuit begins to perform abnormally in the laboratory.

The subsequent waveforms in the signal chain are easier to describe and the details related to them will be discussed below. The Class D waveform is the *voltage* waveform normally measured by probing the axoplasm near the pedicel of a given photoreceptor cell. The Class E waveform actually consists of a closely related family of waveforms defined by a two dimensional matrix. The first set of Class E waveforms are those found at either input terminal, emitter or base, of the Activa of any 1st lateral cell or

any 1st bipolar cell. This is actually a family of waveforms varying in absolute signal amplitude and DC bias voltage. They vary in amplitude and bias because of the voltage divider networks formed by the impedance of the synapse, the associated neurite input impedance and the electrostenolytic supplies involved at these Activa terminals. The second set of very similar waveforms, except for an additional time delay, are those found at the input terminals of the Activa of a 2nd lateral cell or a ganglion cell. In both of these sets, there are slight differences in the delay of the waveform following an impulse of excitation to the photoreceptor cell due to the travel times of the signals along the different path lengths to a given cell.

The Class F waveform is also subdivided into two sub-waveforms primarily for reasons of convenience. This waveform class is fundamentally different from the earlier waveforms. It is associated with the first "action potential" encountered in the signal path. The first variant of the Class F waveform is found experimentally at the output terminal of the ganglion Activa. This location is described as the axoplasm of the ganglion cell in morphological terms. The second variant of the Class F waveform is found experimentally in the plasma at the distal end of each interneuron associated with a ganglion cell. The difference between these waveforms is almost entirely, in the absence of disease, one of time delay. The signal is progressively later in time relative to the excitation illumination due primarily to the regeneration delay associated with each Node of Ranvier.

The Class G waveform is defined primarily for completeness. It is the proposed electrical waveform recovered by the decoding process performed by each of the stellate cells of at the end of the signal projection function in the visual system. This is also a family of closely related waveforms. If the group is studied in detail, they illustrate the differences in temporal, or frequency, response of the various signaling channels using the projection function.

One other waveform category will be of interest after defining the above individual waveforms. That is the composite waveform exemplified by the ERG or frequently the LERG. The characteristics of this waveform are highly variable because it attempts to describe the signaling functions within the retina, or complete eye, *in toto*. This category of waveforms can only be interpreted based on precise knowledge and definition of the illumination and state variables involved.

16.5.1 Collected impulse waveforms of vision

While either the responses due to an impulse or to a rectangular pulse can be used to illustrate the overall operation of the visual system, the impulse responses involve the least overlap. They will be used in this discussion.

Figure 16.5.1-2 summarizes graphically the major waveforms found in animal vision. Each waveform has characteristics unique to itself. The figure shows all of these waveforms based on the same time scale to provide continuity. The following figure will show the binary pulse waveforms in more detail at a different time scale. In this figure, all of the amplitude scales are linear except for the one in the upper left. This scale is different to stress the characteristics of this waveform. The amplitude scale in the lower right is also significantly different. The signal level is typically a factor of 1000 lower for ERG waveforms than for direct probe waveforms.

The waveform of Class C exhibits two unique properties relative to this set. Because of the significant size of the capacitor in the collector circuit, the waveform is that of a Type 2 physical system. It is a continuous curve at the point where the amplitude deviates from the baseline. Furthermore, because of the drop in the voltage of the collector during operation, the electrostenolytic supply must recharge the capacitor following excitation. This recharging current is the reason for the reversal in current after the excitation phase of the waveform.

[The ERG panel shows a curve, originally labeled 250K, taken from Smith & Lamb for a human. By projecting their rising waveform back to the horizontal axis, the time delay expected by this theory is obtained. The initial delay after the initial flash is approximately 3.36 milliseconds using this approach. This corresponds well with the theoretical value from the P/D equation.]

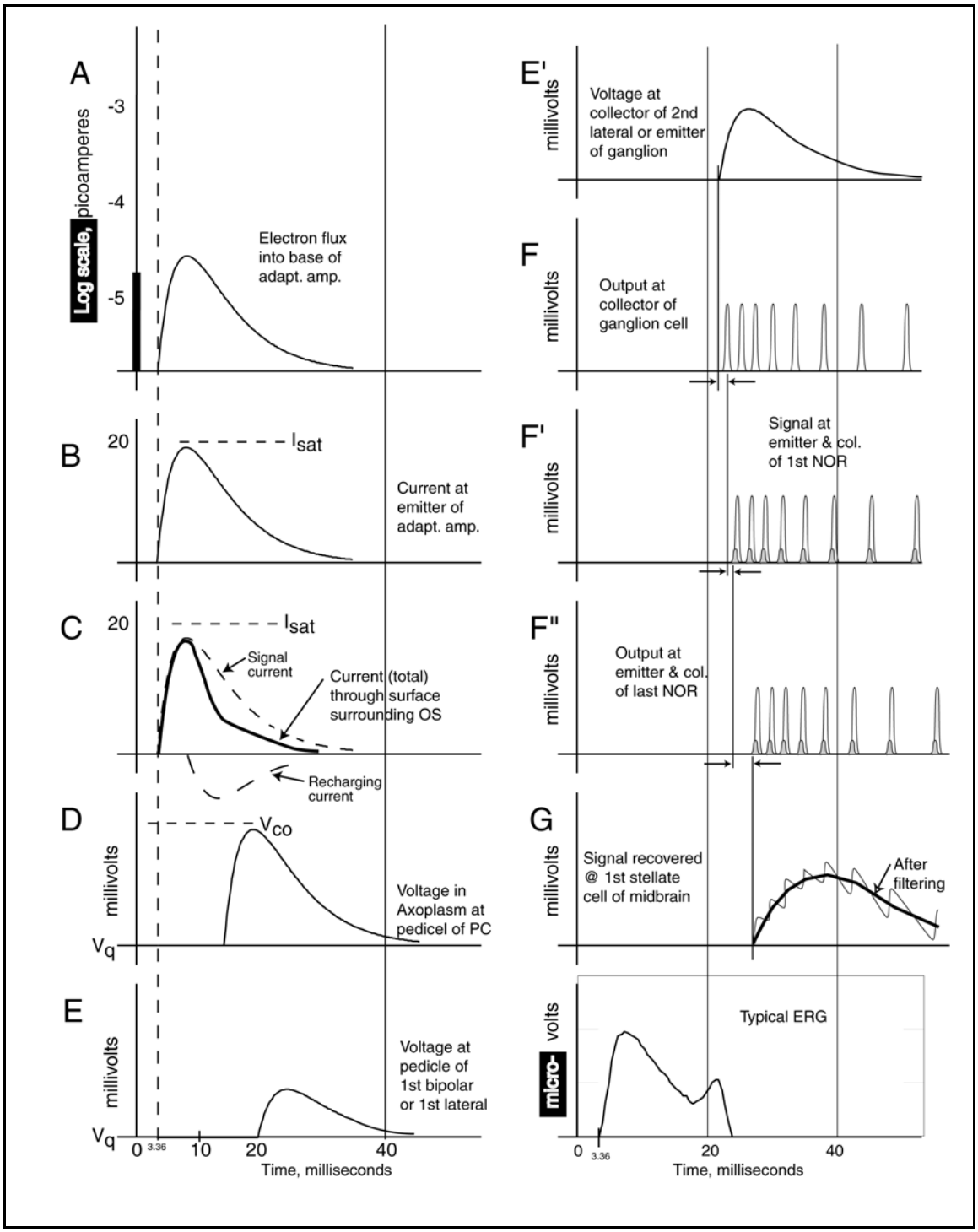


Figure 16.5.1-2 Collected waveform Classes for large signals under pulse excitation as predicted by this work. The time of recovery of waveform (G) in the midbrain is typical. The earliest reported recovery of a (G) waveform in the cerebral cortex is 35 ms. See Text.

16.5.1.1 Waveforms up to the stellate cells of the midbrain

16.5.1.2 Waveforms found within the commissure and extraction engines

16.5.2 Cumulative time delays in vision

Because of the many non-linearities in the visual system, some care must be used in determining the cumulative time delay up to a given point in the system. The problem is complicated in the laboratory when illumination is used that excites more than one of the chromatic photoreceptor channels. The time delay of the individual channels may not be the same. Because of the likelyhood of two different fundamental signal paths within the brain, it is also prudent to calculate the time delay for each path separately. A variety of different types of time delays are involved. Some of these are more difficult to calculate than others and some require an arbitrary standard of measurement. Whereas the P/D equation exhibits a time delay represented by a simple continuous exponential function, the ganglion cell employs a threshold. The time to exceed this threshold depends on the rate of growth of the signal waveform. The projection neurons incorporate a well defined delay. However, the net delay consists of a variety of individual delays. The delay associated with the projection stage decoding process is the most awkward to define. The definition requires a standard of quality be invoked.

In general, all of the circuits between the adaptation amplifiers and the ganglion cells exhibit wide band characteristics that do not constrain the signal passing along the signal path. The time delay of these circuits is generally negligible compared to the time delay of the P/D equation, the adaptation amplifiers and the projection stages. However, the bandwidth of an individual adaptation amplifier may be different than that in adjacent chromatic channels. This can complicate measurements employing broad spectral band illumination. It can also lead to perceived chromatic information that appears to conflict with reality. Beginning with the ganglion cells, the system employs sampled data techniques that have a complicated frequency transfer characteristic and associated time delay.

The time delay related to the P/D equation is independent of the state of the adaptation amplifier. However, it is a function of the level of input signal flux. The time delay associated with the threshold in the ganglion cells is a direct function of the state of the adaptation amplifier and can be a complex function of the input illumination when one or more step changes are made in the illumination level within the time constants of the adaptation amplifier.

The absolute time delay due to the distance between the ganglion cells and the point of cognition is generally the dominant delay in psychophysical experiments. To this delay must be added the delay associated with the response of the animal following cognition. In cases where the animal is at a very low biological temperature, the delay associated with the P/D equation may be significant or even dominant.

In most experiments, only the relative time delay between various conditions is recorded or analyzed. Under small and medium signal conditions, the time delay related to the individual signal paths proximal to, and including the adaptation amplifier, are essentially constant. The dominant variable in the relative time delay equation is the P/D equation.

16.5.2.1 Time delays up to the midbrain

16.5.2.2 Time delays in the alarm channels

The time delay in the alarm channels, assuming the decoding circuits of the brain are arranged like the example in **[Figure 11.7.2-2]**, is simpler to calculate than that in the cognitive channels. This delay does not include the projection stage decoding process. The total delay in these channels, up to the initial stellate cell of the brain, is easily calculated for equal flux illumination and steady state small signal operation. The bulk of the delay is then given by the sum of the:

+ P/D equation delay

+ the adaptation amplifier delay

+ the time required for the signal to exceed the threshold of the ganglion cell

+ the time for the ganglion cell to generate an initial action potential after the threshold is exceeded, and

+ the delay due to the distance between the excited ganglion cell(s) of the retina and the access point of the signal path at the brain

There may be an additional time delay if the signal is received at the LGN and regenerated prior to further transmission over a significant distance within the brain to the point of cognition.

16.5.2.3 Time delays in the cognitive channels

The time delay in the cognitive channels, again assuming the circuits are arranged as in **[Figure 11.7.2-2]**, are essentially the same as listed for the alarm channels except for the addition of the time delay associated with the projection stage decoding circuit at the entry to the cortical system. This time delay is the result of a sampled data process and is difficult to define without careful calculation using the precise parameters of the samplers. The problem is particularly difficult in the chrominance channels where time delay is asymmetrical. The required mathematical background is available in Truxal and later volumes dealing with sampled data systems.

16.5.3 Example of cumulative time delay

The description of the time delay is more complicated for the chrominance channel,and other differencing channels, than for the luminance channel. They will be described separately.

16.5.3.1 A cumulative time delay for luminance

To limit the scope of the discussion, several ground rules will be used initially. It will be assumed that only one chromatic channel is excited and that it is excited by a relatively low level signal from a dark adapted condition. A background illumination introduces other factors that will be discussed later. It will be assumed that this signal is large enough to occupy most of the dynamic range of the signal manipulation portion of the luminance channel without causing a change in the gain of the adaptation amplifier or significant saturation in the logarithmic conversion process at the output of the photoreceptor cell. **Figure 16.5.1-3** highlights the waveforms and time delays found in the visual process under these conditions. Several waveforms are shown with the same labels as defined in Section 11.1.3. (A) represents the response of the photodetection process to a long pulse of radiation meeting the criteria listed above. Among the literature for all animals, the time delay represented by the two Activas in the photoreceptor cell appear to be negligible. Under these conditions and assumptions, this mono phase waveform can be used to describe the Class A, B & D waveforms associated with the photoreceptor cell. It does not accurately describe the Class C waveform which is biphase in character. The waveform is the solution to the P/D equation under the above conditions. It normally exhibits an overshoot at its leading edge under these conditions. There also may be a significant difference in the time constant of the trailing edge of that overshoot and the inherent decay constant,T_R, of the process shown on the right. The process exhibits two distinct time delays related to charge transport and charge density in the excited energy band of the chromophore. The delay described by W is a function of how long it takes charges created within the liquid crystalline lattice of the chromophore to reach the the de-excitation point at the dendritic junctions within the furrows of the disks. It is a function of the rate of incident photon flux and the temperature of the lattice. This time delay is a continuous variable. The waveform exhibits a discontinuity where the waveform departs from the horizontal axis, indicative of a first order dynamic situation and therefore a first order differential equation. It is this delay that authors have traditionally struggled with. They have generally sought to equate this delay with the fixed delay of a multi stage RC filter. The result has been great difficulty in changing the number of stages in the filter as a function of illumination and temperature used in the experiment. The fact that a multi stage RC filter is represented by a higher order differential equation that is continuous where it departs from zero has generally been ignored (figure 2B of Fuortes & Hodgkin for example, is explicit)[60]. Because of the steady state charge density and velocity established in the lattice after only 300 millisecond, the time delay represented by U is different than that described by W. It is a clearing time before the cessation of irradiation impacts the de-excitation process. The vertical scale represents the voltage at the pedicel of the photoreceptor cell relative to its dark adapted value.

---

[60]Fuortes, M. & Hodgkin, A. (1964) Changes in time scale and sensitivity in the ommatidia of *Limulus*. J. Physiol. Vol. 172, pp. 239-263

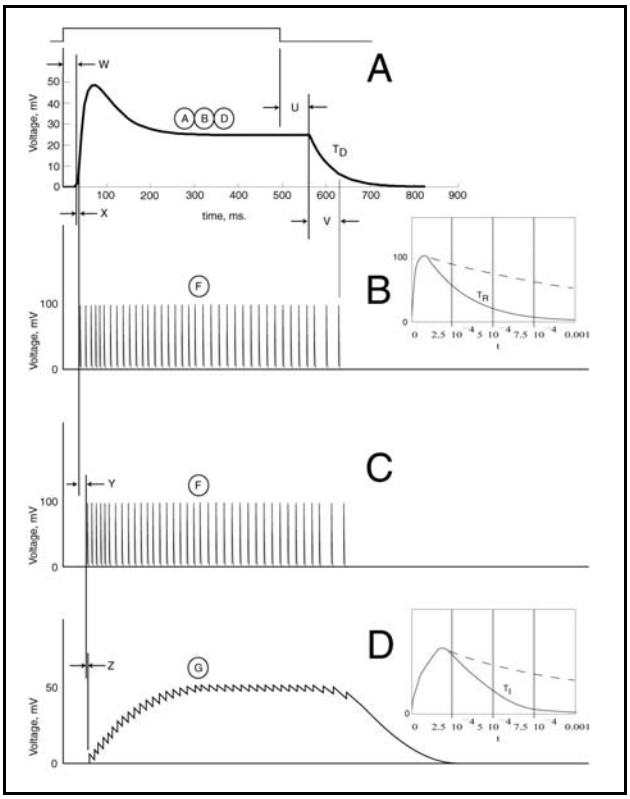


Figure 16.5.1-3 Fundamental waveforms and delay times associated with the visual process. See text for details.

(B) represents the waveform at the input to the first Node of Ranvier of the parasol ganglion cell in this luminance channel. This node is located within the retina. The actual neuron extends out through the optic nerve to the entrance to the visual cortex. It is difficult to show both the overall situation and the details at the same scale. The inset on the right shows the theoretical shape of the action potential derived from the measured waveforms of DeVoe. His work with the warm blooded rat is discussed elsewhere in this work. In contrast to the Class A through D waveforms which are continuous, the waveform consists of two exponential segments with a transition near the peak of the waveform. The exact time of the transition is a very sensitive function of the circuit and Activa parameters. The waveform is asymmetrical and this is frequently seen in photographs of action potential waveforms in the literature. The left part of the figure shows the time between adjacent action potential waveforms varying. It is this variation that encodes the amplitude information concerning the illumination. In this presentation, the fourth and fifth pulses are closest together. The last few pulses are farthest apart. The pulse stream ceases altogether when the input waveform falls below 10 mV. The first pulse is shown occurring after a brief time interval, X, related to the threshold of the ganglion circuit. In the figure, it is assumed to be 10 millivolts for illustration purposes. It is normally much smaller than this, especially in humans. The rise time constant of the first pulse is about 12 microseconds. The decay time is 250 microseconds and the transition is at 75-90 microseconds. The values for rat in-vitro are different base on the work of Bowe, et. al[61]. Unfortunately, their paper does not specify the temperature of there neurons. A temperature of 20-23 Celsius has been assumed.

(C) appears to be identical to (B). However, it is delayed by 5 msec. This figure is meant to represent the same pulse train after many regenerations at Nodes of Ranvier. The time delay is typical of that in a warm blooded animal with an optic nerve nominally 22 mm. long. A neuron of such length can be expected to include 20-30 Nodes of Ranvier, each introducing a delay of about 0.19 msec. The action potential pulses, and their spacing, are assumed to be faithfully reproduced by regenerator circuits having the same nominal time constants as indicated above for the Activa within the soma of the neuron.

(D) illustrates the expected reconstruction of the original waveform at the first stellate cell at the entrance to the cognitive portion of the visual cortex.. The peak of the first sawtooth step is shown delayed by an interval Z which is closely related to the total duration of the received action potential, the ratio of the current capacity of the Activa within the cell and the capacitance of the collector circuit, and the decay time constant of the collector circuit. The individual sawtooth waveforms are shown on the right as they would appear as a voltage across the integrating capacitor of the nominal circuit. The rising part represents the current through the Activa of the stellate cell. It closely approximates the integral of the action potential waveform modified by the presence of a finite collector load impedance. The product of this load

[61]Bowe, C. Kocsis, J. & Waxman, S. (1985) Differences between mammalian ventral and dorsal spinal roots in response to blockade of potassium channels during maturation. Proc. R. Soc. Lond. B vol. 224, pp. 355-366

impedance and the integrating capacitor has a decay time constant given by T_I. Although difficult to see at this scale, the leading edge of each sawtooth is separated from the following leading edge by the same time as in the previous action potential pulse trains. By smoothing this waveform further, the original signal related to the input radiation flux. can be retrieved. The delay relative to the start of the irradiation will be less than 100 msec. depending on the criteria used to specify the delay, and the exact form of the integrators used to recover the signal. In D, a relatively short time constant, T_I, has led to a poor reproduction of the leading edge of the recovered irradiation pulse shape. However, it has minimized the length of the recovered pulse after the cessation of irradiation. This is a tradeoff in the quality of the recovery circuit that can be improved through the use of more sophisticated decoding circuits. Only careful experimental measurements can provide additional data on the quality of the actual circuits used in the brain.

Describing the total delay time under the above ground rules is seen to be complex. The total delay is the sum of four significant delays and a number of minor delays. The major delays, labeled W through Z are of different origins. The initial delay, represented by W, is a transport delay due to a closed loop excitation/de-excitation process described by a differential equation. The second delay, represented by X, is a threshold delay, the time required for a waveform with a finite risetime to exceed a threshold. The third delay, represented by Y, is a simple transport delay. However, this delay is dominated by the regeneration delay associated with multiple Activa circuits and not by the transport of charge along a delay line. The final delay, represented by Z, is basically that of a simple integrator responding to a train of variable spacing action pulses. Under the conditions described above, all of these delays are, to the first order, fixed. The initial delay is however quite susceptible to the temperature of the chromophore lattice, more so than most biological processes.

16.5.3.1.1 Variations from the baseline

With the above baseline available, it is easier to examine additional variants. If the Outer Segment of the same photoreceptor cell has been illuminated for a period of time (which itself involves a transient) prior to the start of a test illumination, the excitation energy band of the chromophore lattice already has bound electrons in it. The presence of these excitons influences both the delay time and the rise time of the initial transient response to illumination. The prior illumination leads to a shorter initial delay and a faster rise time. If the prior illumination continues during the test illumination, the exciton level becomes even higher, as does the amplitude of the signal waveforms. If the prior illumination continues after the cessation of the test illumination, it has little effect on the decay characteristic except to constrain the final value of the function.

In many laboratory investigations, a continuous background illumination is used that excites other photoreceptor cells than that discussed above. Each of the different chromatic channels leading to the same luminance summing circuit must be examined independently. The signal voltages and waveforms, of Class D, associated with each channel will be summed in a summing circuit before proceeding to the Activa of the bipolar cell and then the parasol ganglion cell. Under small signal conditions, the delay and rise time of each of these individual input waveforms will be different. The resulting summed waveform will be quite complex and may exhibit several inflection points along its rising edge, and probably its falling edge. The Class F waveform at the output of the ganglion cell will look much the same as in the figure. However the details of the action pulse spacing may be more complex.. This is not necessarily so because of the integrative aspects of the generator circuit within the ganglion cell.

As the irradiance level is raised, the adaptation amplifier begins to decrease its gain. This tends to reduce the peak amplitude of the Class B waveform passed to the distribution amplifier within the photoreceptor cell. If the signal is quite high, the adaptation amplifier may go into saturation, further limiting the peak amplitude of the Class B waveform. A similar situation occurs in the distribution amplifier. At higher signal amplitudes, the logarithmic aspects of the current to voltage conversion limits the maximum voltage at the pedicel. If the signal is very high, the distribution amplifier will actually be cut off and the output voltage will reach the intrinsic electrostenolytic membrane potential for the cell. Determining exactly which of these limitation comes into play when requires further laboratory investigation, probably on individual species.

A few words are needed concerning the Class C waveform frequently measure in the laboratory by probing the space between Outer Segments. There are two important considerations. Generally, the probe senses the short term average of the currents through the IPM impedance from a number of individual cells, not all of the same spectral sensitivity. Thus the resulting waveform is similar to that at the summing network between a group of photoreceptor cells and a bipolar cell forming the luminance signal. For larger amplitude irradiances however, the Class C waveform differs from the Class D waveform because it is not

logarithmically converted to a voltage. Furthermore, the Class C waveform is bipolar, it represents the current passing through the collector capacitor of the adaptation amplifier due its discharge by the adaptation amplifier as well as the current passing through the electrostenolytic process in order to recharge that capacitor.

Equations can be written for all of the above variants. However, because of the non-linearities involved, their value is greater for intellectual insight than for experimental design purposes.

16.5.3.2 A cumulative time delay for chrominance

With the above description of the luminance channel as background, it is easier to explore the time delays of the chrominance, and other differencing, channels. The key to understanding the operation of these channels is the understanding of the differencing circuit of the lateral cells. There are two separate conditions under which the cell remains quiescent and there is no signal output. The eye can be in complete darkness, resulting in no signals input to the circuit. Alternately, the signals input to the circuit can be of equal amplitude, resulting in no net signal on the emitter of the Activa. Both situations result in an achromatic perception by the subject. To obtain a chromatic perception, it is necessary for the photoreceptors to generate a net signal at the emitter. Whereas it is easy to determine the initial time delay for each chromatic channel, the effective time delay of the difference depends on the relative amplitude of the two signals. This difference depends on a variety of parameters; the luminance of the source (filtered through the physical optics) relative to the absorption band of each chromophore, the state of adaptation of each adaptation amplifier, and the total number of input synapses of the lateral cell associated with each chromophoric channel. Using much the same ground rules as for the luminance channel, the two chromatic signals of Class D are likely to exhibit nearly the same initial time delay and negligible time delay is associated with the lateral cells. However, the signal presented to the midget ganglion cell of the chrominance channel is biphasic. The midget cell must accomodate this waveform. To do this, the ganglion cell is biased into a range where it produces a continuous series of action potentials with a nominal period of 33 ms. Because of this bias, the midget ganglion cell cannot employ an input threshold and the delay related to such a process, labeled X, is zero. The information is encoded by changing the time interval between the pulses as above. However, the encoding now either increases or decreases the nominal period between the pulses. The primary difference between the difference channels and the luminance channels involves the recovery of this encoded information. This difference is illustrated in **Figure 16.5.1-4** for two saturated chromatic conditions. The integrator is assumed to be perfect with an infinite decay time constant. The upper waveform represents the recovered signal for a pulse stream of 10 ms. period (nominally 100 Hz frequency). For purposes of continuity, the time delay before any response of this circuit is shown as the sum of the earlier time delays, labeled in the previous figure. The lower waveform represents the recovery of a 100ms. period (nominally 10 Hz.) pulse stream. The time delay associated with the decoding process, based on any threshold criteria, is strikingly different. For intermediate signal saturation conditions, the time constant varies monotonically between these two extremes. This theoretical asymmetry and the well documented asymmetry in the human chromatic visual response leads to the conclusion that the chromatic photoreceptor signals are not randomly connected to either the inverting or non-inverting inputs of the chromatic difference channels. Every experiment leads to the conclusion that the M-channel photoreceptors connect to the lateral cells so as to lead to the shorter period, higher frequency, pulse trains from the midget ganglion cells.

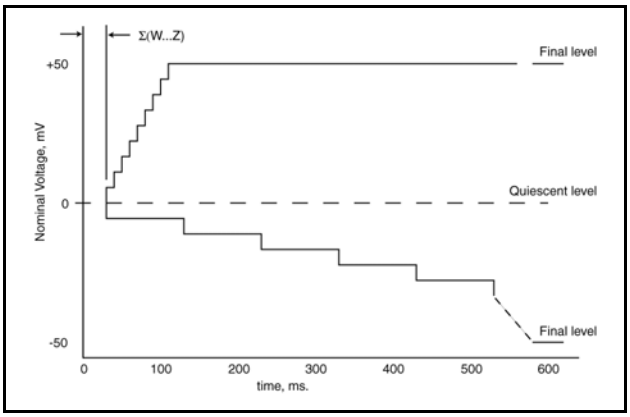


Figure 16.5.1-4 Asymmetrical operation of the signal projection codec. XXX XXX

16.6 Evaluation and Interpretation via Psychophysiological Experiments

The range of psychophysiological experiments in the literature, involving humans and in some cases primates, is truly immense. Many of these experiments are

"XXX" because they fail to define and/or control all of the variables relevant to the overall experiment. However, some of the data may be interpretable if re-evaluated within the context of a complete model. This section will attempt to review some of the data in the context of the model proposed here. Initially, relatively simple experiments will be reviewed as they provide the highest potential for rationalization.

The dynamic range of the human eye is so great, it is difficult to perform experiments characterizing the entire range. Most experiments on human subjects have concentrated on two nominal conditions based on the assumption that they are independent and unique conditions; the photopic and scotopic luminance conditions. The C.I.E. has promulgated Standards based on this situation.

Frequently, a subject has not been able to exhibit sufficiently high sensitivity to small signals. In this case, more than one size field has been employed. The C.I.E. provides luminous efficiency functions for both 2 and 10 degree fields.

16.6.1 Intermediate complexity situations

16.6.1.1 The Complete Luminosity Function

The C.I.E. recognized Standard Luminous Efficiency Functions, both photopic and scotopic, are empirically derived. They are based on nominal states of adaptation, relatively wide spectral bandwidth excitation, and the averaging of the results of a number of investigators employing a relatively small number of subjects. The results ostensibly represent the brightness sensitivity of the eye and are not affected by the hue of the stimuli. However, the difficulty of separating these two sensations is widely recognized.

Because the results are entirely subjective, there has been no need to define a term for the output function of this human psychophysical process. To provide a more theoretical foundation for this process, it is desirable to define the transfer function, *T*, and the response function, *R*, where *R* is essentially equal to the product of the stimulus and the transfer function, *T*, of the human visual system. By equating the response function, *R*, to the threshold sensitivity of the eye under the associated conditions, it is possible to calculate, *T*, which corresponds to the luminosity function of the human visual process. This luminosity function remains a function of the background adaptation level, the width of the spectral filters used to evaluate the subjects, as well as other parameters to be discussed below. Thus, to accurately define these functions, the investigator must do so after prescribing these variables and parameters. Once these values are specified, it is possible to compute a Luminous Efficiency Function at a given stimulus level; two well known, but imprecisely defined, levels being the photopic level and the scotopic level. The Luminous Efficiency Function is then the luminosity function, *T*, evaluated at the prescribed parameters and normalized in terms of its peak value.

[[what stimuli units should be used see W & S pp. 250-253 about transition of candela from a broadband based source to a monochromatic source]]

Based on this model, it is possible to define the luminosity function of human vision in terms of its two fundamental variables and other basic parameters. The complete human luminosity function is a function of both the illuminance, E_v, applied to the subject, the color temperature of the illuminance, T, the wavelength, λ, and the width, $\Delta\lambda$, of the spectral filter used to establish the wavelength. It is given by the equation:

[[terminology is in a mess, see W & S pp. 250-260]]

$K(E_v,T, \lambda,\Delta\lambda) =$ Eq: 16.6.1-1

where K is known as the luminous efficacy for monochromatic radiant flux P sub (e, λ) of wavelength λ

By evaluating the above function at different illuminances, it is possible to account for four different luminous response functions, the hypertopic, photopic, mezotopic and scotopic. It is then clear that these different functions cannot be emulated by manipulation of two or more of them to obtain the third. It is also clear that each function can be obtained without requiring the interaction of any of the chromophoric channels.

By evaluating the mezotopic function at various illuminances, it is seen that the Purkinje Shift is not a smooth one with a peak passing through intermediate wavelengths. It is created by a saddle with the peak sensitivity switching from that associated with the XXX to that associated with the XXX.

These results are contrary to much of the conventional wisdom concerning the mezotopic sensitivity curve as summarized in Davson[62].

16.6.1.2 The Dark Adaptation Function

[see van den Brink in Vision Res vol. 2 for best data and Stevens & Stevens fig 10 for concept. **See Section 17.6.1**]]
Figure XXX indicates the scope of the problem of defining the dark and light adaptation functions for the animal eye. One must account for the transient performance of:

+ the iris
+ the P/D equation
+ the adaptor stage
+ the vascular system

Most of these elements exhibit performance that classifies them as "state variables"; performance is a function of their state at the start of the test and the change in signal due to the test. Proper test documentation and control is complicated in the case of state variables. Figure XXX provides some data relevant to this situation.

16.6.2 Complex situations

Some of the more complex test situations involve the determination of the frequency passband of the visual system because of the number of variables involved. Frequency response tests can be configured in a variety of ways;

+ Simple temporal tests involving a flashing light source

+ Simple spatial tests involving a moving light source

+ Spatial tests involving a moving grating

+ Spatial tests involving a grating with elements changing "color" in a periodic manner

+ Flicker tests using .....

16.6.2.1 Evaluation of flicker tests

The evaluation of the frequency response of the visual system is difficult because of the number of variables involved and the number of different signal paths that may be involved. A particularly awkward problem observed in the literature is the practice of measuring the illumination level used with an instrument calibrated for a given spectrum, usually photopic, and a given color temperature source. Values are then reported based on this instrument even if the color temperature is different or if the spectral response of the visual system changes drastically with illumination level. Little recognition is given to the fact that the signals at the output of the photoreceptors may have changed drastically in character and the signals in the differencing paths may have also changed drastically.

It is important that fundamental flicker tests be performed with spectral colors, preferably corresponding to wavelengths that only affect one chromophoric signaling channel. Once this type of data is understood, it is possible to interpret test data resulting from broad band testing much more accurately. This is especially true if the test involves highly structured input images such as gratings.

---

[62]Davson, H. Pg. 442

16.6.2.1.1 Flicker testing with broadband illumination

[see laming on right of wallboard or in Seeing Contour and Colour, Kulikowski , note change in frequency passband with Troland level.]]

16.6.3 Evaluation using Fourier Transform spaces

The mathematical manipulations involving the Fourier transform and various convolution integrals have taken on a greater importance in vision research in recent times. However, the level of mathematical rigor employed by many investigators has frequently been less than optimum. A particularly awkward practice has been the presentation of photographic records of the spatial frequency plane using a black background and no axes. Without axes, the investigators frequently overlook the relationships associated with the two orthogonal axes.

Many investigators have recognized Moire Effects in many of their data sets but not always discussed their origin and impact in sufficient detail to be clearly understood.

This section will address some of these problems and provide a clearer interpretation of some of the figures found in the literature.

16.6.3.1 The general Fourier Transform

There are many books available describing the Fourier Transform, at many levels of completeness[63], the volume by Wilson is particularly applicable to optics[64]. Unfortunately, they all require a knowledge of the calculus to interpret adequately and are designed to be used with a teacher providing the practical details orally.

16.6.3.1.1 The one-dimensional Fourier Transform

As the name implies, Fourier Transforms occur in pairs. An equation describing the amplitude (and phase if appropriate) of a function along an axis can be transformed into an equivalent expression describing the frequency components of the same function. The complete description of both pairs must be over the integral limits of plus and minus infinity, and the expressions generally involve complex notation that defines both an amplitude and a phase component.

---

[63]James, J. (1995) A Student's Guide to Fourier Transforms. Cambridge: Cambridge Univ. Press

[64]Wilson, R. (1995) Fourier Series and Optical Transform Techniques in Contemporary Optics. NY: Wiley & Sons.

The Fourier transform $F(y)$ of the function $f(x)$ is defined by the integral

$$F(y) = \int_{-\infty}^{+\infty} f(x) \exp(-2\pi jxy)\, dx$$

The function $f(x)$ can be deduced from $F(y)$ by the inverse Fourier transform,

$$f(x) = \int_{-\infty}^{+\infty} F(y) \exp(2\pi jxy)\, dy$$

When x represents time, y is the frequency.

The functions f(x) and F(y) can take on any form. There are a variety of simplifications that can be employed when dealing with Fourier Transforms under specialized conditions. The Fast Fourier Transform is a common one. It suffers from limited accuracy but, as its name implies, is much faster to compute using a digital computer. Another simplification is the LaPlace Transform. It only requires integration over the interval from 0 to plus infinity when the functions exhibit certain symmetries. Large tables of both Fourier and LaPlace transform pairs are readily available.

16.6.3.1.2 The two-dimensional Fourier Transform

While the two-dimensional Fourier Transform is more formidable on paper, the extension of the basic idea is simple and direct.

$$A(p,q) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} F(x,y) e^{2\pi i(px+qy)} dxdy$$

$$F(x,y) = \int_{q=-\infty}^{\infty} \int_{p=-\infty}^{\infty} A(p,q) e^{-2\pi i(px+qy)} dpdq$$

A(p,q) is the function in the frequency plane. F(x,y) is the function in the spatial plane. x & y are coordinates in the spatial plane, p & q are coordinates in the frequency plane.

16.6.3.1.3 The optical solution to the Fourier Transform

Although computation of a two-dimensional Fourier Transform is often formidable for something as complex as an optical image, there is a very simple way around the problem. **Figure 16.6.1-1** shows the simple process of obtaining the Fourier Transform of a two-dimensional pattern using optical methods. The technique relies upon the fact that the images in the two focal planes of a lens are Fourier Transforms of each other. While the transform in the image plane is a complex pattern in amplitude and phase, the

complete optical transform function (OTF), it is common to disregard the optical phase function (OPF) and only deal with the amplitude portion of the complete function generally known as the modulation transfer function (MTF). However, this procedure can introduce errors that are not obvious to the inexperienced investigator.

Practical examples of this process will appear in the following sections.

16.6.3.2 Parameters related to Moire, sampling and correlation

Before proceeding to discuss the material in **Section 15.2.5.2** of this work, it is important to review the relationships between the image and the process of sampling such an image.

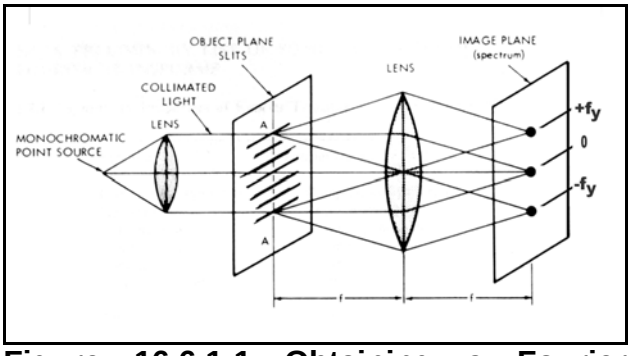


Figure 16.6.1-1 Obtaining a Fourier Transform by optical means. The one-dimensional form shown is directly extendable to two-dimensions using a more complex image in the object plane.

Williams paraphrased Bracewell[65] with regard to the Moire Effect and MacLeod, et. al[66]. discussed a visual nonlinearity in vision relying upon some of the concepts of Moire, sampling and correlation. It appears a broader discussion of these phenomenon and mechanisms is needed.

First, it should be noted that the Moire Effect is totally independent of any sampling process. The remainder of this section will focus on the Moire Effect. The following section will discuss sampling.

The Moire Effect is explained within the context of a linear process based primarily on the superposition principle. If two similar wave functions (not necessarily sinewaves) are caused to be summed (in one or two dimensions), the resulting signal(s) will exhibit relative maximums and relative minimums periodically that are not correlated with the other maximums and minimums. These values are explained in terms of the Moire Effect. The resulting patterns exhibit frequencies given by the sum and difference of the two original frequencies. If the waveforms are two dimensional, the sum and difference can be taken with regard to the axis of one of the original waveforms. Alternately, the sum and difference can be taken of the components of the wave resolved upon any two orthogonal axes. This is the approach used generally to plot the spatial frequency components in Fourier Space (as in the MacLeod paper). In the MacLeod case, the vertical frequency component is described as the product of the frequency of either of the incident waveforms multiplied by the cosine of half the angle between the two fringe patterns. The horizontal component is described in terms of the frequency of the incident waveforms multiplied by the sine of the angle between the two fringe patterns.

16.6.3.3 Sampling at the basic level

Sampling is a more complex mechanism of determining the amplitude of a signal over a generally short interval, of time or space, in order to reconstruct a copy of the signal at a different time or place. The process generally employs multiplication and is therefore mathematically nonlinear.

When an image is sampled, generally for the purpose of converting the image into electrical signals that can be transmitted to a remote location, it is important to interpret the situation carefully. In general, the sampling process does not involve any Moire Effect or other artifact. However, there are a set of guidelines on how well the signal can be reconstructed. These guidelines are generally attributed to Nyquist. Their application to audio and code based communication (where the phase of the recovered signal is not significant) is generally attributed to Shannon. Nyquist determined two critical frequencies. Here they will be called the Nyquist frequency and the Nyquist (ambiguity) limit. Working with signal waveforms that last for an indefinite period of time, Nyquist said the sampling frequency must exceed the highest frequency component in the signal if the potential recovery is to be perfect. For signals of shorter duration, the quality of the recovered signal is a function of the phase between the original signal and the sampling mechanism. For a window of finite dimension, the quality of the recovered signal begins to degrade when the sampling frequency is less than twice the relevant frequency component in the original signal. This criteria defines the Nyquist ambiguity limit.

> The difference between the Nyquist frequency and the Nyquist ambiguity limit and the significance of the sampling window within which the signal is sampled, have not generally been appreciated in the visual community.

The sampling process generates a signal that is devoid of any Moire or other interference patterns. It is the reconstruction process that introduces these effects. The observed effect is due to the same superposition principle as discussed above. It should be noted that Moire patterns are not a required artifact in image reconstruction. They are routinely avoided in high performance image transmission equipment. The patterns observed in black and white television are the result of economic decisions. If the shape of the electron beam used to recreate the picture on a cathode ray tube is given a cosine-squared shape, and the scan lines are spaced so that the adjacent beams overlap appropriately, a uniformly illuminated screen can

---

[65]Bracewell, R. (1978) [new 2000 edition available] The Fourier Transform and Its Applications. NY: McGraw-Hill

[66]MacLeod, D. Williams, D. & Makous, W. (1992) A visual nonlinearity fed by single cones. Vision Res. vol. 32, no. 2, pp 347-363

be obtained without any scanning structure. If sampled data is introduced to form an image, the picture can be reconstructed without any Moire pattern related to the direction of the scanning pattern.

> As a side-note, the interlaced scan used in conventional television was introduced to reduce the Moire Effect in the reconstructed black and white pictures without incurring a significant cost. When color television was introduced, Moire Effects became a major problem. A large multi-corporate-laboratory study was performed to find methods to minimize the effect of Moire patterns in the resulting color pictures. A “dot-sequential” technique was adopted to minimize this problem.

16.6.3.4 Sampling and image reconstruction

In the process of sampling, it is interesting to note where artifacts may enter the data stream(s). In the case of vision, each individual photoreceptor samples a small area of the image presented to the retina. In the absence of significant scattering of the light, each sample is entirely independent of the sample associated with all other photoreceptors. Because of motion between the photoreceptor and the image, due to tremor, the sampling generates a continuous signal with respect to time that describes the image as it is sampled. In the case of the foveola, each of these signals is transmitted to the midbrain over an independent neural circuit. Thus, each circuit is devoid of any information related in any way to any other channel. The investigations of Nyquist and others play no role in these signals with respect to the spatial frequency domain of the image. However, these circuits are band limited by the adaptation amplifiers of the individual photoreceptors and are sampled in time as they are converted into action potentials for projection in the stage 3 circuits of vision.

Overlooking this bandlimiting and temporal sampling for the moment, the individual signals from the foveola arrive at the pretectum for processing within a multi-dimensional correlator. It is in this correlator that the samples taken from the original scene are analyzed. It is here that correlation between the spatial frequency content of the image and the spatial frequency content associated with the sampler (the retinal mosaic) are first observed. The impact of the sampling process can not be observed at the retina. When observed in the correlator, the characteristics of the observed image are determined by a host of parameters, including both the spatial pitch of the original photoreceptors and the spatial correlation interval of the correlator. It is also determined by the parameters of the stage 2 signal processing (for the foveola, this is largely limited to the adaptation amplifier) and the stage 3 signal projection circuits.

Although the process of interest is 2-dimensional, the sampling mechanism can be explored initially as a one-dimensional process. Let it be assumed that the 2-dimensional sampling array is orthogonal and regular (points at the center of each cell of a checkerboard). Looking at the signals along only one of the major axes of this array formed by the application of an image, several features can be noted.

The Nyquist frequency of a 1-dimensional sampling process is defined as the reciprocal of twice the sampler spacing (cell spacing of the checkerboard along one axis). If the image contains spatial frequency components at precisely the Nyquist sampling frequency (or multiples of it), the recovered signal will not contain these frequency components, only a DC component representing their presence.

There are two forms of aliasing, or incorrect reproduction of the data, involved in sampling at non-Nyquist frequencies using a less than infinitely long duration window. First, up to a frequency defined as the Nyquist frequency, all of the information contained within the original image can be recovered. However, at frequencies above one-half the Nyquist frequency, the amplitude of the recovered signal will be sensitive to the phase of the spatial frequency component of the input signal. This lower frequency is frequently called the Nyquist (ambiguity) limit. Errors related to spatial frequencies between the Nyquist frequency and the Nyquist limit will be called (Nyquist) phase errors here. Second, for spatial frequency components in the scene that are above the Nyquist frequency, a frequency component will appear in the recovered signal given by the original component frequency minus the Nyquist frequency. This effect is usually described in terms of frequency-fold-back.. The frequency observed under these conditions will be called a (Nyquist) frequency error. Here again the phase relationship between the input signal and the sampling signal plays a role.

These relationships are shown in **Figure 16.6.1-2**. The figure is a caricature and not drawn to reflect the actual mathematical relationships relative to amplitude of the reconstructed signal. Frame (a) shows the ambiguity in the amplitude of the recovered frequency component as a function of the phase between the original signal and the sampling mechanism. The ambiguity begins at one-half the Nyquist frequency. At the Nyquist frequency, the amplitude of the recovered frequency component is indeterminate. The point where the ambiguity begins is sometimes called the Nyquist (ambiguity) limit. The wide variation in the reported component amplitude is due to the duration of the original signal component within the sampling window.

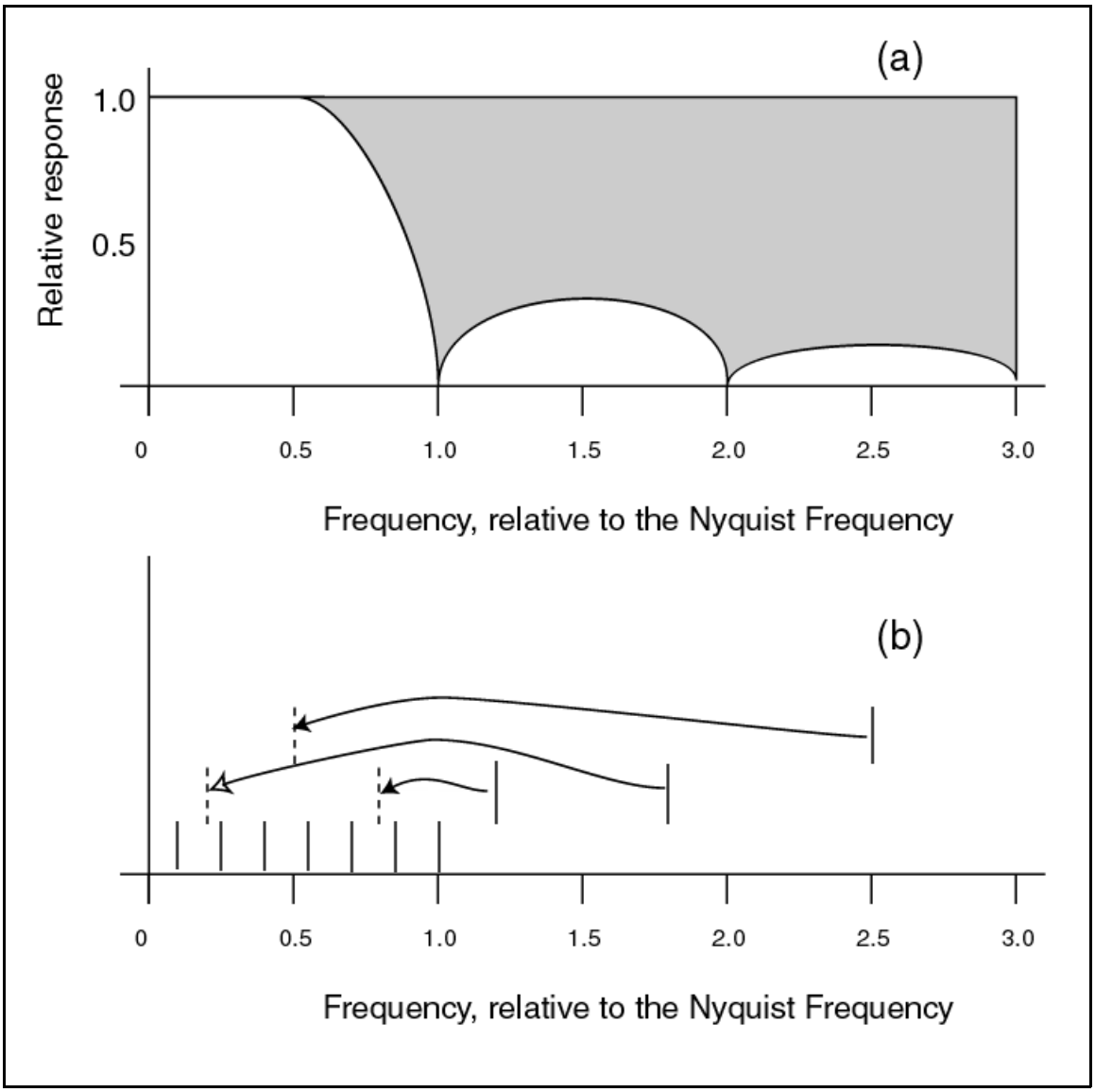


Figure 16.6.1-2 Relationship between Nyquist frequency and Nyquist limit. (a) shows the ambiguity in the amplitude of a reconstructed component following sampling. The ambiguity (sometimes called the Nyquist limit) begins at one-half the Nyquist frequency. (b) shows the fold-back of frequencies above the Nyquist frequency (and above twice the Nyquist frequency). As a frequency component in the original signal approaches twice the Nyquist frequency, it is reproduced at a frequency approaching zero.

Text book solutions of the under-sampling problem usually assume the frequency component in the signal is infinite in duration. Under this condition, there is no ambiguity in the amplitude of the frequency components of the output signal. The recovered amplitude is 100% of the original signal amplitude. However, if the frequency component is limited in duration (by a physical aperture), the amplitude and frequency of the recovered component is a function of the phase between the signal and the sampling pattern within the window. This is why contrast thresholds are found to be a function of the window used with the test signal. It is also why conventional three-bar test targets (which are constrained to a square

window by design) frequently appear to only have two bars near threshold (maximum resolvable frequency).

Frame (b) illustrates the frequency folding encountered when a signal is under-sampled. Frequency components in the original signal occurring above the Nyquist frequency will be reproduced at a lower frequency reflecting the difference infrequency between the component and the Nyquist frequency (or its harmonic). The result can be a very low frequency component not present in the original signal. The zebra stripes, to be discussed below, are a representation of this mechanism.

The Moire patterns formed in two-dimensional examples of the above sampling mechanism and phenomenon are widely used in the printing, wind-tunnel testing and strength-of-materials fields. The pitch between the straight Moire lines is an accurate gauge of the difference in frequency between the samples (between two pieces of window screen as an example). The waviness is an accurate portrayal of the phase error along the lines. The discussion of the mechanism underlying the Moire Effect is an awkward one. The Moire effect is technically an effect inherent in the superposition process (an inherently linear process). The Moire effect is represented by a linear and algebraic difference in frequency between the two generating spatial frequency patterns. However, in many situations, the signals also participate in a non linear process that introduces permanent frequency components related to those observed in the superposition process.

16.6.3.5 Application of Fourier and Moire concepts in vision

16.6.3.5.1 The Fourier frequency space for crossed fringe

MacLeod, et. al. presented the Fourier frequency plane of two interferometrically generated fringes with an angle between their major axes. Their figure can be interpreted using the concepts and figure in **Section 16.6.3.1.3**. The original caption to their figure suggested that the spatial plane, shown on the left contained no horizontal frequency component (although one was clearly present) and justified the statement using a white on black print of the Fourier frequency plane. **Figure 16.6.1-3** reproduces the figure with the polarity of the print reversed and axes added. The original figure consists of the pair of dots at the top, the pair of dots at the bottom and the dot at the center of the figure. The addition of the axes makes the separation of the image into vertical and horizontal components obvious. The horizontal component (which is real) is described by the horizontal spacing between the two dots at the top of the image. It should be noted that Fourier Transforms exhibit both positive and negative frequency components, related to the integration interval. The negative components can frequently be ignored. However, in this case, the axes have been chosen to pass between the two dot patterns for convenience and the negative frequencies in the horizontal plane are meaningful. In this example, the two fringe patterns are of the same frequency, F. The horizontal component is given by $2 \cdot F \sin\Theta$ using the vertical axis as a reference. The vertical component is given by $F \cdot \cos\Theta$. Although not shown clearly, the variation in the brightness of the vertical pattern on the left is twice as large as the variation in horizontal brightness because of the summation of the two $F \cdot \cos\Theta$ components. This summation is based on the fact the two fringe patterns are of equal brightness but not temporally coherent. In this case, their energy is added algebraically. If the two fringes had originated from the same laser source and were within the temporal coherence distance of the laser, the two fringes would be added vectorially. As a result, the brightness of the interference pattern would be a function of the temporal phase between the two fringes. While not noticeable in this example, the dots indicating the frequency of the fringes, and the dot at the center, in the frequency plane have a finite diameter that is a function of the size of the window used to frame the pattern on the left. This feature will become more obvious in later examples.

Note the linearity present in the above figure. If either of the fringes is removed, the remaining pattern on the left will be at a frequency of F at an angle of Θ with respect to the vertical. There will be a horizontal component related to F but there will not be a difference component. The frequency pattern on the right will consist of only one pair of dots at an angle of Θ and symmetrically located along an axis through the center dot representing the window.

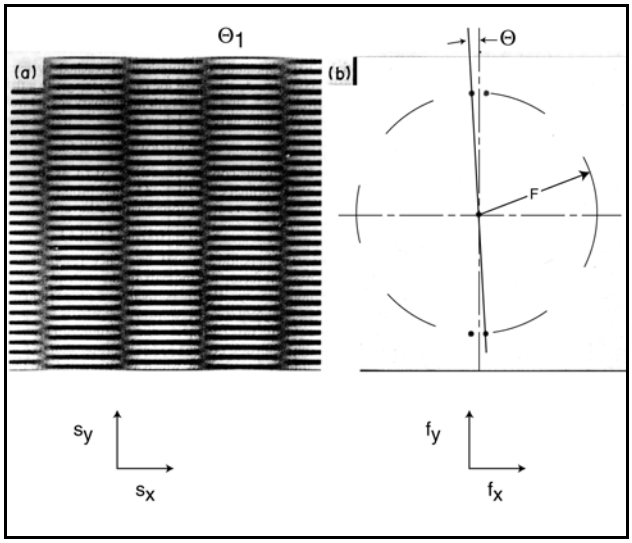


Figure 16.6.1-3 Fourier planes of two interference patterns of the same spatial frequency superimposed at an angle. The vertical frequency component in (a) is F•cosΘ where Q is relative to the vertical axis. The horizontal frequency component in (a) is given by 2•FsinΘ. The patterns are the result of a linear superposition. The intensity variation in the vertical component is twice that of the horizontal component. Modified example from Macleod, et. al. 1992.

16.6.3.5.2 The Fourier frequency space for crossed fringes applied to a nonlinear mechanism

MacLeod, et. al. also represented, by example, the result of passing the interference pattern shown above through a nonlinear mechanism. This process is the same as that used in demodulating a radio signal. The result is the generation of new components in the output that are independent of each other. The process is illustrated in **Figure 16.6.1-4**

This pattern is significantly different from that above. The nonlinear mechanism has generated a unique new pair of components at the frequency of G = ±(2•FsinQ). The magnitude of these frequency components can be confirmed by drawing the parallelopipeds as shown. The significant difference is that if this signal is now filtered to remove the components at the frequency F, the components at G will be unaffected. Although not illustrated in the example, the nonlinear mechanism also generated new vertical components as well, at a frequency, H = ±(2•FcosQ). These components are outside the graph shown. If the two fringes had not been at the same frequency, F, the resulting sum and difference frequencies would have been more complicated and a pair of frequencies would appear along the vertical axis near zero. This would have caused a coarse frequency component in the vertical direction in the frame on the left similar to that shown in the horizontal direction.

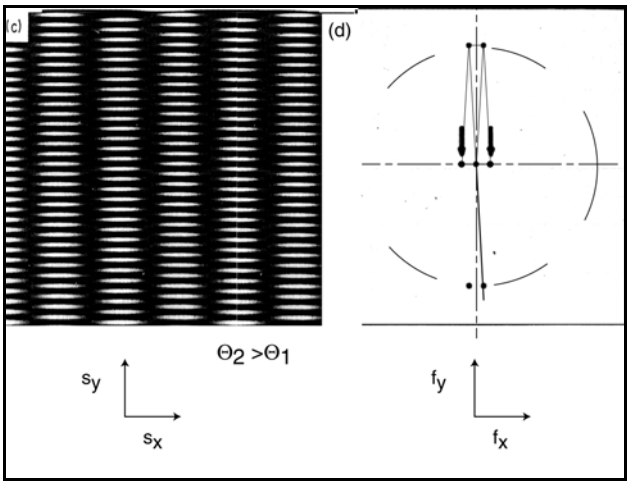


Figure 16.6.1-4 Fourier planes of the sum of two interfering fringe patterns after passing through a nonlinear mechanism.

16.6.3.5.3 The Fourier frequency space for the foveola

There has been a difference of opinion in the literature as to whether any Moire pattern, observed in psychophysical experiments, is a function of the orderliness of the retinal mosaic or not. Yellott has said that the disorder in the retinal mosaic largely prevents aliasing[67]. Miller & Bernard take an opposing view, that the primary reason for the small amount of aliasing observed is due to the size of the individual photoreceptors and the limiting resolution of the optical system due to Rayleigh diffraction[68]. However, the Miller & Bernard premise is largely conceptually based and appears flawed. It is based on an earlier paper by Snyder & Miller[69]. It also refers to the inner segments of a retina rather than the actual measured data of the outer segments in Yellott. Miller & Bernard stress their position that the inner segment is the critical element in the photosensing task. In a separate calculation, Snyder & Miller use an absorption coefficient of ε = 10%. This number appears to be very low. A value nearer ε =80% is suggested based on the overall database in the literature.

Miller & Bernard take an unorthodox view of the sampling process, attributing the more basic mechanism to the Nyquist ambiguity limit than to the Nyquist sampling frequency. This leads to a variety of

---

[67]Yellott, J. (1982) Spectral analysis of sampling by photoreceptors: topological disorder prevents aliasing. *Vision Res*. vol. 22, pp 1205-1210

[68] Miller, W. & Bernard, G. (1983) Averaging over the foveal receptor aperture curtails aliasing. *Vision Res*. vol. 23, no. 12, pp 1365-1369

[69]Snyder, A. & Miller, W. (1977) Photoreceptor diameter and spacing for highest resolving power. J. Opt. Soc. Am. vol. 67, no. 5, pp 696-698

difficulties. As an example, their equation (2) should not be limited to $\nu<3\nu_N$. Frequency foldback occurs at any frequency that the system is capable of transmitting. They did define the Nyquist ambiguity limit clearly. However, many subsequent readers have confused this value with the Nyquist sampling frequency, that is two times higher.

The more basic problem in their discussion is that it assumes a regular repeating close-packed hexagonal array for the retina (see **Section 3.2.2.3** for an alternate view). They chose to refer to only one of the three principle axes of such an array as illustrated in Snyder & Miller. **Figure 16.6.1-5** is a composite of data from both Yellott and Snyder & Miller. While Snyder & Miller chose to speak in terms of the horizontal pitch between staggered rows in the close packed array resulting in their use of $\Delta\varphi = \lambda/(D\bullet(3)^{0.5})$ for the center-to-center spacing of the photoreceptors, the vertical pitch between staggered rows is even smaller. The center-to-center spacing without stagger give larger values for $\Delta\varphi$. The question is, which formulation of the spacing is best in a discussion related to aliasing? The left frame is reproduced in Yellott from Polyak. It shows a 12' by 13' area of the foveola (about 25 photoreceptors in diameter). The middle frame from Yellott based on the left frame seems to make the choice moot. Yellott's Fourier Transform of Polyak's micrograph (which does not involve the real physiological optics of the eye) is nearly circular and shows no preferred axis. This is very far from the situation proposed by the Miller group. There is a vast array of data on the Fourier transform of a regular closed packed hexagonal crystal. The transform of such an array shows a very clear set of axes. Except for the ellipticity of the transform, the real retina used by Yellott shows no preferred axes.

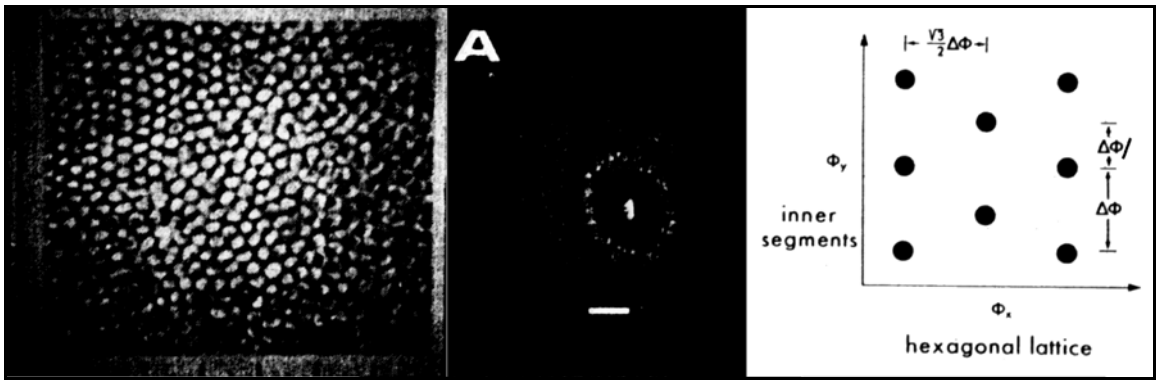


Figure 16.6.1-5 Comparison of the data of Polyak and the analyses of Yellott and Snyder & Miller. See references for better copies.

Considerably more detail is found in the original Yellott Fourier transform. (The unusual shape of the central spot was not explained. It may be due to the reproduction process.)

The horizontal bar at the bottom is described as indicating a distance of 100 cycles/degree. This primary ring is considerably below the 150 cycles/degree or 245 cycles/degree limit defined in Miller & Snyder. The locus of the first order ellipse fitting this ring has a major to minor diameter close to 1.155:1 at an angle of about 120 degrees from the horizontal. The ratio of 1.155:1 is important as it is the reciprocal of the ratio 0.866:1. This is the ratio of the horizontal photoreceptor pitch of Snyder & Miller to the full photoreceptor-to-photoreceptor pitch. This full receptor pitch occurs along axes at 60, 90 and 120 degrees from the horizontal axes. No information was provided by Polyak concerning the orientation of the figure relative to physiological vertical.

Of probably more importance is the frequency supplied by the transform and the deviation in that frequency. By careful measurement, the fundamental frequency of the retinal array is given as 108 cycles/degree along the short axis and 128 cycles/degree along the long axis. The ratio between these frequencies would be 1.155:1 if either of these values numbers were changed by only 2.6%.

The 2-sigma deviation of the points in the main ring of the Fourier frequency plane the nominal ellipse is less than ±10%. There is more information in the original published figure. There are a considerable number of points forming a broad secondary halo between the primary locus and an ellipse of 150% the size of the primary locus. Whether this halo is an artifact or not cannot be determined.

This analyses provides important insights. The Nyquist frequency (not limit) of the human eye is a function of the orientation of the retina. Along one axis in the only sample available, it is approximately 1.155:1

higher than along the other major axes of a close-packed hexagonal array. This ratio is equivalent to 1:(3)$^{0.5}$/2. The Nyquist frequency for the normal human eye is approximately 108-110 cycles/degree except along this preferred axis where it is 125-128 cycles/degree. Further experimentation is needed to confirm whether this higher frequency is coincident with the horizontal axis of the eye. If it is, it may have significant implications regarding how text characters are formed using vertical and horizontal strokes.

It can also be seen that the average photoreceptor-to-photoreceptor spacing in the human eye is equivalent to a spacing of approximately 108-128 cycles/degree. This corresponds to the Nyquist frequency (not limit) for the array. For an eye with a focal length of 16.68 mm in air, the center-to-center spacing would calculate, from this data, as 1.45 microns at 100 cycles/degree. For the preferred axis, the center-to-center spacing would calculate to 1.2 microns at 125 cycles/degree. These measured value are considerably below the three microns assumed by the Miller papers. However, they are in general agreement with the measured values presented by Polyak for the diameter of individual photoreceptors (1.0 to 1.3 microns at the center and edge of the foveola of a human)[70]. A more precise diameter for the photoreceptors could be obtained if the focal length of the subjects eye were known.

Using a test set such as the flash photomicrograph test bench of Roorda & Williams, it is possible to take spatial images and Fourier Transform images of eyes *in-vivo*. It is also possible to measure the focal length of these eyes *in-vivo*. The performance of these tests would provide the actual Nyquist frequency of the subject eyes along multiple axes, ala Yellott, and the precise center-to-center spacing of the photoreceptors calculated from th measured focal lengths, *in-vivo*. By varying the size and location of the patch of the retina evaluated, the impact of these factors could also be evaluated.

It appears the general suppression of any Moire or sampling artifacts is due to the disarray in the photoreceptors of the retinal mosaic, even in the foveola where it is considered most uniform. Whether the physiological optics contributes to the further suppression of these secondary effects appears academic.

16.6.3.3 Measurements comparing M– and L–channel contrasts

Several authors have provided data comparing the M– and L–channel contrasts in attempts to evaluate the luminance and iso-luminance performance of the eye. Most of this data has been collected using inadequate color temperature light sources. This has invariably led to the ignoring of the S–channel, and the resulting P–channel, contrast performance in these investigations. The investigators have also usually assumed the L–channel spectrum peaks near 560-570 nm. As a result, the measurements are usually of low signal-to-noise ratio and frequently show lower contrast performance in the L–channel than in the

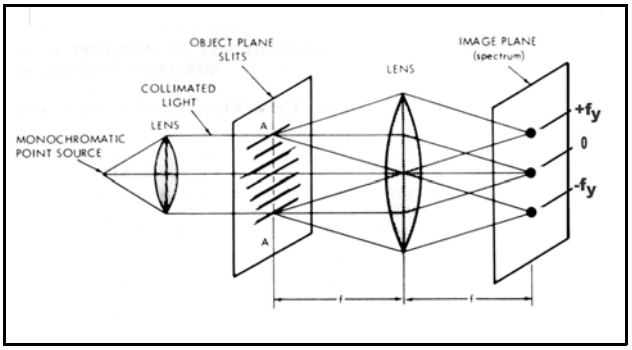


Figure 16.6.1-6 Obtaining a Fourier Transform by optical means. The one-dimensional form shown is directly extendable to two-dimensions using a more complex image in the object plane.

[70]Polyak, S. (1957) The Vertebrate Visual System. Chicago, Il: Univ. Chicago Press pp 267

M–channel. Recent tests of this type were reported by Gegenfurtner & Hawken[71]. The tests are usually designed to investigate the signal processing channels of the cortex (see **Section 15.2.5**).

A serious need exists to repeat some of the experiments in this area using a light source with a color temperature near 7053 K, spectral filters optimized to match the actual spectral channels of vision, and insuring the light level used is well within the photopic range and defined in the protocol.

16.7 Evaluation and Interpretation via electrophysiological Experiments

[xxx Figure numbers wrong because of some spacing positioning problems. Figures missing.]

There are a number of pieces of data (and collections of data) that are worthy of further review here because of the insight they provide as the operation of the visual system in animals. In order to accomplish this objective, three steps will be taken;

+ a discussion will first be provided of the operating scenario to be evaluated

+ a nomograph related to that scenario will be developed based on the overall model

+ data from the literature (in as close to original form as possible) will be placed on the nomograph

+ the key features present in the data will be discussed from the perspective of the model.

Many of the parameters involved in the development and explanation of these scenarios are multidimensional, and this fact must be accepted and explored in order to gain a useful understanding of the overall effect on the visual process. Specific examples will appear shortly.

[Be sure and review the work of Armington and others in the ERG folder as part of this section, probably later, possibly by combining 16.6.2 & 16.6.3 under one heading and then showing how they differentiate.]

16.7.1 Evaluation and Interpretation of Ganglion Cell Output

The work of Palmer, Barlow, Levick, Straschill, and (waveform with pre-emphasis in pulse train) have provided very useful data on the output of the ganglion cells in response to luminous stimuli. To interpret this data in detail, it is necessary to have a complete circuit diagram of the relevant portions of the visual system available. These portions include the circuits reaching Stage 3 via the luminance, chrominance, and appearance paths (as well as the polarization paths in some insects). In developing the portions of the model relating to these prototypical paths, several features have been found which suggest the elaboration of certain circuit elements or circuit types not explicitly mentioned in the literature. In general, these elaborations include;

+ the reiteration that the type of modulation[72] employed by the ganglion cells is “time delay interval”, a form of phase modulation (not to be confused with the simpler frequency modulation). This is important because of the subtle details shown in Levick’s graphs in later sections.

+ the separation of the luminance path from the related chrominance, appearance (and possibly the polarization) paths based on signal characteristics; where the luminance signals are assumed to be *monophasic* and passed through the signal chain from the adaptor circuit of the photoreceptor cell to the input of the ganglion cell *without signal inversion.*

+ the grouping of the chrominance, appearance and polarization paths under the general heading of difference paths; where the difference signal is clearly biphasic and it creation necessarily involves the inversion of at least one of the luminance signals at the input to a lateral cell.

---

[71]Gegenfurtner, K. & Hawken, M. (1996) Interaction of motion and color in the visual pathways *TINS* vol 19, pp 394-401

[72]There are subtle matters of definition between the encoding of a analog signal in a pulse domain and the modulation of a nominal carrier frequency by an analog signal that should be addressed by any serious investigator. Levick in the Handbook, vol VII/3 pp. 588-89 gives some references to the literature but they speak with a lack of clarity in this area; Levick & Zacks speak of deliberate concealment by coding ?!.

+ the recognition of the fact that the signal undergoing inversion at the input to the lateral cell may also experience signal amplification in the same circuit.

+ the differentiation of the ganglion cells into two classes, those exhibiting a maintained discharge in the form of an action potential and those that do not. A tentative relationship has been assigned to two types of ganglion cells recognized on morphological grounds; the midget ganglion cell as one exhibiting a maintained discharge and responding to a biphasic input signal, and the parasol ganglion cell as one that responds to a monophasic input signal and does not exhibit a maintained discharge.

+ the recognition that at least the ganglion cell responding to a monophasic (luminance) signal may incorporate a pre-emphasis circuit in its input structure.

+ the realization that the bias applied to the ganglion cell can make a significant difference in how the overall circuit performs

This work will utilize the amercine cell as the prototypical lateral cell without meaning to de-emphasize the importance of the horizontal cells; primarily because of a lack of substantive data differentiating the two broad classes of lateral cells, the horizontal cells generally located at the interface of the fiber layer and the outer plexiform layer and the amercine cells generally located within the inner plexiform layer (or at the interface between the inner nuclear layer and the inner plexiform layer--depending on the literary source). Jameson[73] provides a clearer discussion and figure describing the role of these two types of cells than do earlier works but it is still mostly conceptual.

16.7.1.1 The scenario

There is a large amount of data describing the response of the ganglion cell to a monophasic signal resulting from illumination of a luminance type signal path. Much of this data was discussed in earlier chapters at the component level. More recently, data has begun to appear related to the response of a ganglion cell to a biphasic signal, resulting from either differences in the chromatic differencing circuits or in the appearance (spatial luminous) differencing circuits. Although the common experimental regime related to on center/off surround experiments becomes quite complex when both spatial and chromatic parameters become involved--as they inevitably do, they will be kept separate here in conformity with the initial ground rules that only the simplest prototypical signal paths will be explored in order to avoid entanglement in the quagmire of actual neural organization.

The important task of this section is to elucidate the significant elements of both the fundamental luminance signal path and the basic difference signal path. Clearly, it is important when performing experiments involving the difference paths, that the two input signals be as independent as possible of each other. This requires that chromatic signals be chosen with due regard to the actual spectral absorption characteristics of the individual chromophores. This generally means that the use of broadband and colloquially named colors (red, blue, green, etc.) is not acceptable. The use of narrowband filters is indicated, preferably centered on the absorption spectra of the appropriate UV, S-, M- or L-channel. If state of the art data is to be collected, the narrowband filters should have a passband of not more than +/-5 nm.: otherwise, the data may be significantly filtered (smoothed) during acquisition. Similarly, if the L-channel is to be used in the experiments, it is important to recognize that the signal, at least in the OS to PC portion of the circuit, is related to the square of the illumination whereas the other channels are linearly related to the input illumination (prior to the onset of saturation in the OS). With these considerations under control, it is important to point out that the individual luminance channels exhibit independent adaptation characteristics which are related to the adaptor portion of the circuit of the PC neuron. Thus, selective adaptation can have a profound impact on the data. Also of interest, both here and in the psychophysical experiments is the phase sensitivity of the output of the differencing circuits to the input of the luminance channels. This is particularly important when performing spatial frequency experiments involving multicolor stimuli; some experimenters have displayed a dip in the middle region of the spatial frequency response of humans while others have displayed a peak. Due to inadequate experiment design, *the two experimenters were performing different experiments* without knowing it.

The block diagram and circuit diagram appropriate to this discussion have been presented in **Figures**

---

[73]Jameson, D. (1982) Handbook of Sensory Physiology vol. VII/4 pg. 382

16.2.1-1 and **14.5.3-1** respectively. No random noise source are shown in these figures. Because of the photoelectric nature of the chromophores (band gap of over 2.0 volts) and the high gain of the adaptation amplifiers, little noise is contributed by the early circuits in the visual signal chain. The predominant noise in the system is due to quantum noise associated with the stimulus and the noise level associated with the stellate (signal recovery) cells within the cortex.

Figure 16.7.1-1 [discriminator characteristic, show only heavy line and change names xxx]] adopted from Hurvich & Jameson[74], seems to isolate the two difference channels of human vision. One involves the S- minus M- signal and the second involves the L- minus M- channel. This is the interpretation of this author and not that of Hurvich & Jameson who are focused on the Hering approach to the Opponent Theory of color vision. They continue to stress a colloquial nomenclature in order to remain focused on that Theory.

Since the data was acquired psychophysically, the data does not completely isolate the chrominance channels. The chromatic valence is a conceptual parameter. Stiles & Birch (1959) have shown (and the theory of this work concurs) that the peak perception of red actually occurs at a wavelength of 700 nm rather than near 610 nm. It is proposed that the true M– L chrominance signal is represented by the heavy line segment from 532 nm to about 610 nm plus a section continuing to rise out to about 700 nm. Similarly, the M– S chrominance signal is represented by the lighter line from 437 nm to 550 nm. As discussed in Section xxx, the cortex changes its reliance on the recovered data from the M–L to the M–S signal channels in the vicinity of 532 nm based primarily on an absolute magnitude criteria. Thus, the values between 437 nm and 532 nm on the heavy line, and 532 to 700 nm on the light line, are not used in perception. Because of this switching, it is not necessary to explain the presence of a high positive chromatic valence (corresponding to red in the concept of Hurvich & Jameson) at short wavelengths.

The nominal zero values in the portions of the two curves used by the cortex occur near the value of 494 and 572 nm defined in this work as the neutral points in the two chrominance channels.

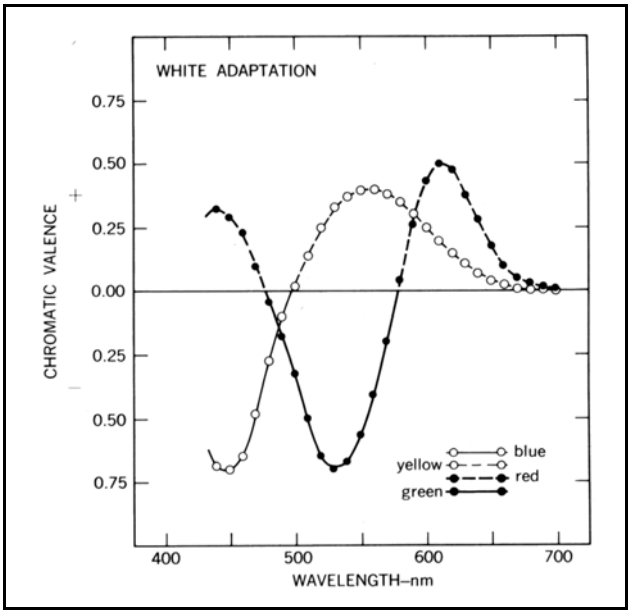


Figure 16.7.1-1 Empirical data supporting the two-chrominance channel concept of this work. The data includes other parameters besides the two chromatic discrimination functions. See text. From Hurvich & Jameson, 1955.

16.7.1.3 The data and its key features

[xxx section still needs editing]
Straschill[75] has presented a series of recordings from the optic tract and the Lateral Geniculate Body of cat. The data only covers an illumination range of four orders of magnitude and the specifics of the photometry used is apparently open to question. Levick discusses these considerations and the fact that the graphs

[74]Hurvich, Jameson, D. in W & S pg 457
[75]Straschill, M. (1966) Title in German (Roughly Neural activity in the optical tract and LGB as a function of illumination intensity) Kybernetik vol. 3, pp. 1-8

presented were not an exhaustive set[76]. Levick also discusses his work in collaboration with Barlow. The important information in Straschill is that a number of axons in the optic tract can be placed in a group that exhibit the same nominal frequency under the apparently same illumination; the term apparent is used since the data in Levick is not sufficiently detailed. Extrapolating the data in **Figure 16.7.1-2** show that the axons of the optic tract, reflecting a higher frequency with illumination, have a maintained frequency of about 20 Hertz at 0.0 log light intensity in lm/m^2 while those exhibiting a decrease in frequency with illumination have a maintained frequency near 36 Hertz at 0.0 log light intensity in lm/m^2 . The data from the LGN are only shown for reference since it is not clear how they relate to the optic tract data, except that they are from the same species of animal and reflect somewhat lower frequencies, roughly 50% of those in the optic tract. The upper right trace may be indicative of additional processing being carried out in the LGN. A unique feature of these traces is the remarkably straight line found on a graph with linear-logarithmic axes. These straight lines are not compatible with a linear relationship between the parameters plotted; they represent a logarithmic relationship. The general form of this relationship is:

$y = K \log x$ **or** $y = (1-w) \log x$ **or** $y = K(1 - w) \log x$ where K = a fixed gain function and w = a feedback coefficient; and x = illumination, y = frequency and the log symbol corresponds to the conversion of a current to a voltage by a diode impedance element.

The last form appears to provide the requisite flexibility to fit the proposed model most appropriately. In a later section, an attempt will be made to evaluate K and w based on the available data .

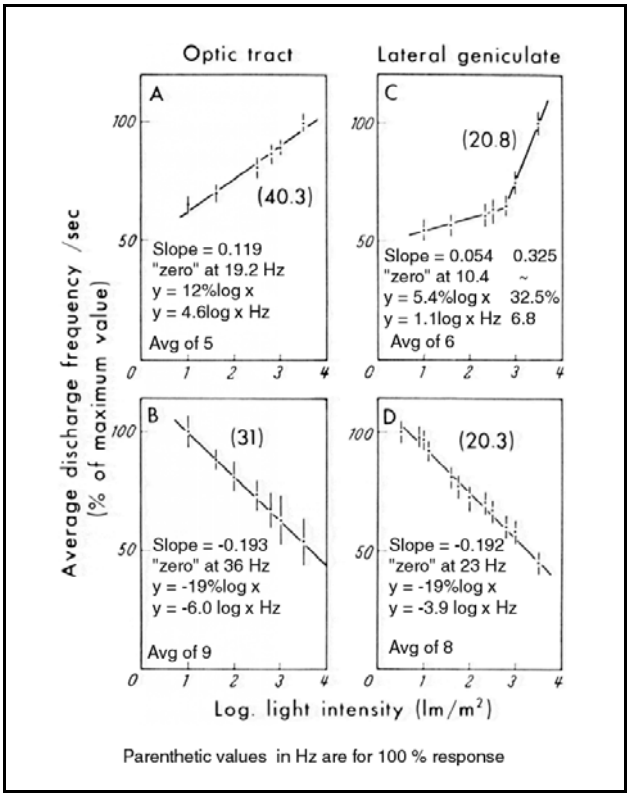


Figure 16.7.1-2 Action potential parameters of cat in optic tract and LGN. Absolute values in parenthesis equal maximum value at 100%. Tabular values by extrapolation from the original data. It is suggested that 10 Lm/m2 approximates 620 cat scotopic trolands. From Straschill, 1966

[76]Levick, W. (1982) Maintained discharges in the visual system and its role for information processing. in Handbook of Sensory Physiology, vol. VII/4 pp. 575-598

Barlow & Levick[77] have presented more comprehensive data for the cat which is presented here as **Figure 16.7.1-3**. In this early exploratory work, they used a common interpretation of their data. They describe the left frame as due to an "On-centre" and the right frame as an "Off-center ganglion cell without defining the effect of any surround (which was receiving the same stimulus. This work takes a different view. It interprets the left frame as representing a typical luminance channel ganglion cell and the right frame as representing a typical chrominance channel ganglion cell. Under this interpretation, the firing rate of each cell is determined differently.

Disregard convergence for the moment, and explore the left frame as representing a fundamental luminance channel. The luminance channel ganglion cell is a driven monopulse oscillator. It is biased so as to respond with a pulse rate proportional to the average potential of the bipolar cell connected with it. This average is directly related to the average potential of the pedicle of the photoreceptor cell connected to this bipolar cell. The pedicle potential represents the specific potential at the output of the adaptation amplifier. Under this interpretation, the mean firing rate of 50 Hz represents the nominal firing rate over the photopic operating range. Within the photopic range, the average gain of the adaptation amplifier is inversely proportional to the stimulus. Modulation of the scene within this range would cause this rate to vary between about 10 and 80 Hz. If the light level exceeds the photopic range, the adaptation amplifier is no longer able to compensate for the stimulus, and the output pulse rate rises with signal level. Conversely, at stimulus levels below the photopic range, the adaptation amplifier cannot provide sufficient gain to compensate for the low stimulus level. The firing rate falls until a minimum rate is reached. This minimum rate is determined by the bias level of the output amplifier of the photoreceptor cell in the absence of any stimulus. The regions discussed above are shown as a dashed overlay to the figure. The ranges are nominal, but consistent with the rest of this work. The photopic range is shown as 3.5 log units in width. The mesotopic range covers about 2 orders of magnitude.

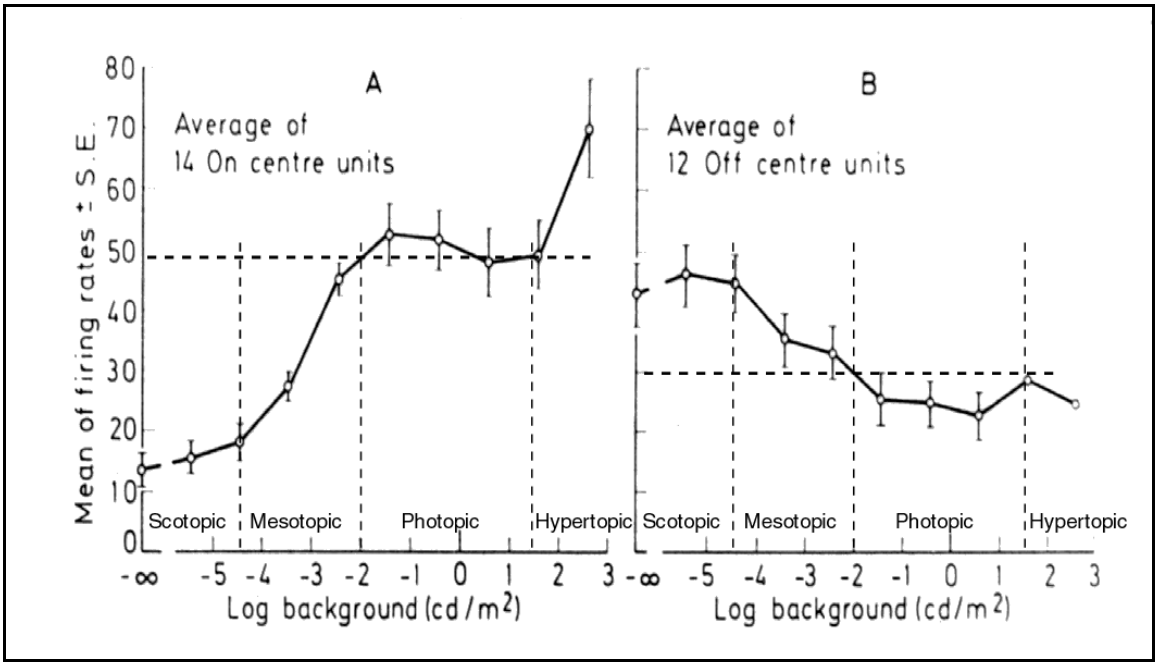


Figure 16.7.1-3 Action potential parameters for cat in the presence of a uniform background field. Statistics for 12 units on left, and 14 units on right. Dashed lines added by this author to show adaptation capability of the eye of cat. Left frame shows a luminance channel ganglion cell. A nominal 50 Hz pulse rate is maintained over the photopic range. Right frame shows a chrominance channel ganglion cell. A nominal 30 Hz rate is maintained over the photopic range. Pupil area, 7 mm². Modified from Barlow & Levick, 1969.

[77]Barlow, H. & Levick (1969) Changes in the maintained discharge with adaptation level in the cat retina. *J. Physiol.* vol. 202, pp 699-718

Again, disregard convergence for the moment, and explore the right frame as representing a fundamental chrominance channel. The chrominance channel ganglion cell is a free-running pulse oscillator. It is biased to produce a series of pulses at a rate near 30 Hz in the absence of any control signal. Modulating this oscillator can result in the pulse rate varying between a nominal zero and 80-100 Hz. This ganglion cell receives its input from a lateral cell. The lateral cell produces a signal that is the difference between the output of two spectrally different photoreceptor cells. The lateral cell output potential can be described as representing either the difference S-M or L-M. In the case at hand, the individual adaptation amplifiers are free to adapt individually in an attempt to provide a nominal perception of white. For this nominal value, the lateral cell output potential is nominally zero and the ganglion cell pulse rate is at the quiescent value of 30 Hz. This is the situation over the photopic range shown. At higher stimulus levels, both of the relevant adaptation amplifiers will reach their minimum gain capability. However, the difference value at the lateral cell output will remain near the quiescent value of 30 Hz. At stimulus levels below the photopic range, the lateral cell will continue to difference the signals from the two spectral channels and the ganglion cell will continue to respond to this difference. However, the adaptation amplifiers in these two channels will reach maximum gain at different absolute levels. As a result, an individual ganglion cell pulse rate will vary in an arbitrary manner. It may rise or fall in this range. The value of the signal provided in this range is not normally accepted by the cortex. The color signal processing within the cortex is suppressed when the average luminance channel pulse rate falls below about 25 Hz.

Under this interpretation, the data in the two frames do not show any center-surround characteristics. However, when convergence is reintroduced, such secondary characteristics may be observed.

This data provides a limiting value for the frequency at zero absolute illumination. The frequencies are in the same range as those of Straschill; 10-15 Hertz for a luminance channel ganglion cell at zero stimulation. Barlow & Levick note that average pulse frequencies below 1 Hz and above 70 Hz are very rare at any adaptation level. It appears that the error bars in Barlow & Levick are due primarily to their test set and the small number of samples in their data set. The color temperature of their light source was not given and they indicate their conversion of photopic and scotopic light levels into flux levels was approximate.

These two waveforms can be placed on the basic nomograph for summing and differencing circuits. The data are presented as dotted lines in [**Figure 16.1.2-7**] xxx using the conversion 1 foot-lambert = 3.53 Cd/m^2 .

As discussed in earlier sections, experimental design is very important in performing electrophysiological tests dealing with the ganglion cells; if the probe used presents too much capacitance to the cell, the characteristics of the pulse train may be affected--either in terms of the inter pulse delay or in terms of the shape of the individual pulses.

As indicated in section 16.1.2, the error bars associated with the data of Barlow and Levick (working with cats) are not necessarily associated with experimental error, they may well be due to systemic causes; i. e., a random noise source at the input to the ganglion cell and the absence of a bias at that input creating a threshold that would prevent that noise from reaching the output. Thus, in at least the cat, sensitivity would be enhanced over that of the human at the expense of a noisy presentation under very low light conditions. There are methods of suppressing this type of noise at the demodulator in Stage 4; they will not be addressed here.

The character of the error bars in Barlow and Levick provide strong corroboration for the method of signal modulation proposed in this model. The character of the data of both XXX and Barlow & Levick strongly support the concept of, and provide a means of calibrating, the Adaptor proposed in this model.

16.7.2 Evaluation and Interpretation of ERG waveforms

Attempts to understand the visual system and to relate various diseases to electrophysiological measurements have been under way since the beginning of the electrical age. The procedural aspects of evaluating the performance of the eye have been standardized for clinical purposes[78]. However, the fundamental mechanisms creating the composite waveforms used by the clinician are not well documented. These fundamentals can now be presented in a comprehensive manner. However, the topology and topography of the visual system is so complex that it is still difficult to discuss the actual situation in a brief account. This Section will explore the subject in fours sub-sections:

+ A comprehensive signal flow and grounding plan
+ A discussion of the fundamentals and complexities inherent in the ERG
+ A similar discusion for the LERG
+ Some pertinent relationships between the basic waveforms and these techniques

16.7.2.1 A comprehensive signal flow, signal delay and grounding plan

To understand the origin of the ERG, it is important to model the multitude of signal sources within the retina. This section will look at both the gross and fine structure of the retina related to signal generation and signal distribution. Signal distribution must be separated into the orthodromic portion used by the visual system and the pathological portion producing signals frequently sensed by man with special instrumentation. The orthodromic portion travels within the neural paths leading to the cortex. The pathological portion propagates from a multitude of individual points within an electrolytic matrix of finite impedance. Most of these signals radiate from their individual point of origin. Signals traveling within the electrolytic matrix travel quite slowly relative to the phasic signals departing from the ganglion cells. The electrotonic signal flow within the stage 1 and stage 2 neural circuits also travel quite slowly. These slow signals are the source of the ERG.

The global electrical topology of the retina has not been studied prior to this work. However, much can be inferred from related studies in the literature.

16.7.2.1.1 The physical arrangement of circuit elements in the retina

The morphologist has long recognized a series of membranes related to the flow of metabolites within the eye. These include, Bruch's membrane, the outer limiting membrane (OLM) and the inner limiting membrane (ILM). However, the purpose of these membranes was not clear. It has also been realized that the RPE consists of cells packed very close together and they might constitute a barrier. Previously, this work has discussed the need to isolate the IPM in order to protect the delicate chromophores, at least during their transit from the RPE to the Outer Segments. This requirement provides a justification for considering the RPE and the OLM as metabolic barriers within the retina. It appears that they also constitute electrical insulators. There is also a need to control the metabolic interface between the retina and the vitreous humor. The inner limiting membrane appears to support this metabolic requirement, creating an electrical insulator at the same time. Cunha-Vaz[79] has edited the papers of a symposia on the metabolic aspects of these barriers. The terminology used there, however, is not directly translatable to the above situation. In general, the RPE is considered the "outer blood-retinal barrier (BRB)." In the portion of the ocular globe covered by the retina, the ILM appears equivalent to their "inner blood-vitreous barrier (BVB)." No discussion of the OLM was found in that text. **Figure 16.7.2-1** (repeats **Figure 11.1.2-3**) shows these surfaces, beginning with the choroid and ending with the vitreous humor. Also shown in the figure is a fundamental signaling path for the visual process. It begins with the RPE cells separating the vascular matrix, supported by the choroid artery, from the IPM. The IPM is in turn separated from the INM by the OLM. Each of these regions is filled with an electrolyte of finite impedance and is connected to the remaining electrolytes of the body through an undefined series of other electrolytes or electrolyte filled chambers. This is indicated symbolically by the resistors and capacitors on the left. Because of the electrolytic nature of the medium, the transport velocity of charge through this medium is very slow compared to the velocity in metallic conductors. As a result, these networks act as delay lines.

---

[78]Marmor, M. Arden, G. Nilsson, S-E. :& Zrenner, E. (writing as the International Standardization Committee) (1991) Standard for clinical electroretinography. *In* Heckenlively & Arden *ed*., Principles and practice of clinical electrophysiology of vision. St. Louis, MO: Mosby Year Book, pp. 283-288

[79]Cunha-Vaz, J. (1979) The blood-retinal barriers NY: Plenum Press

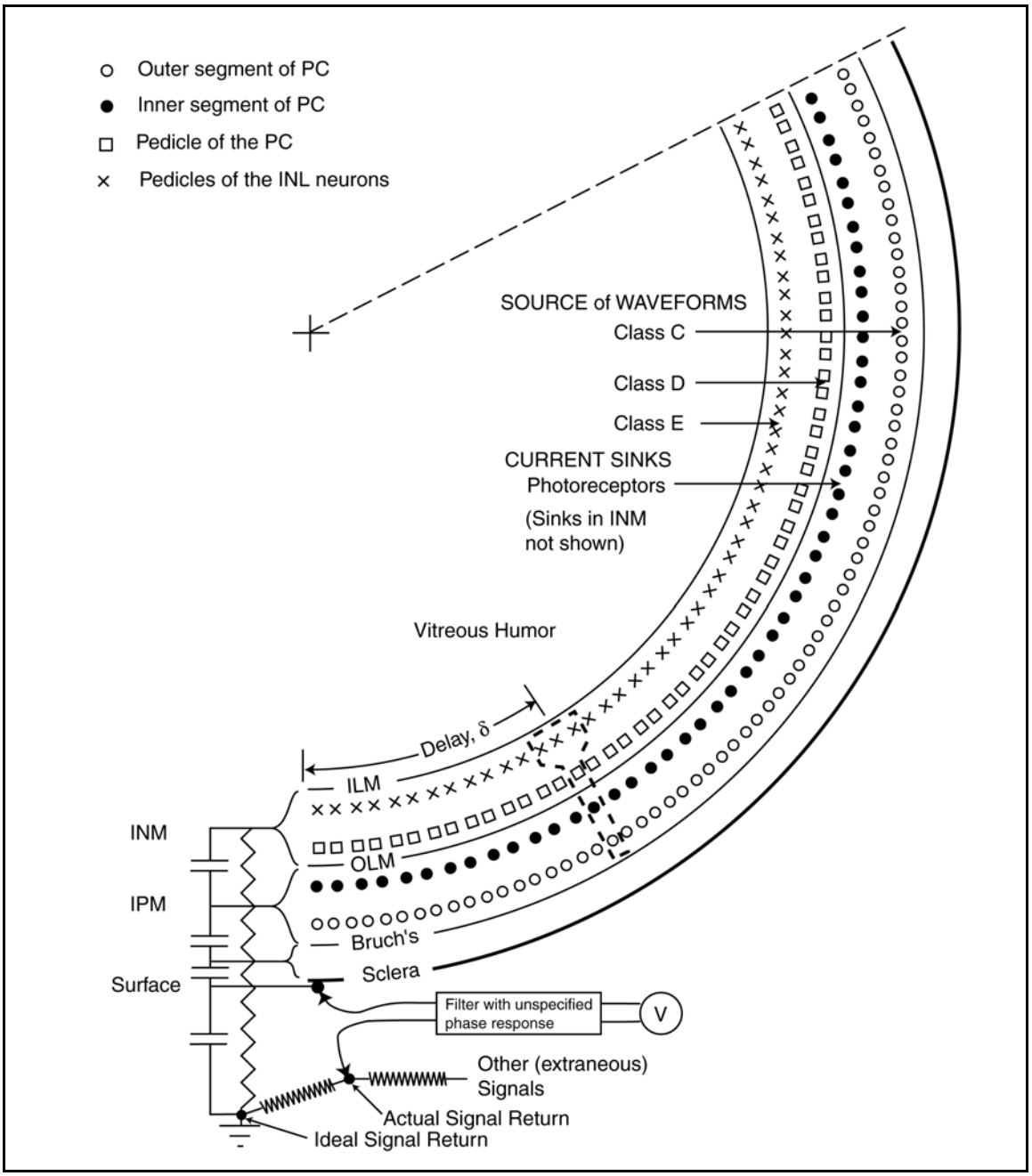


Figure 16.7.2-1 **The signal and grounding plan of the human retina. ERG lead placement relative to the topography of the retina. The source of the different classes of waveforms are shown on the right in this cross-sectional view of the retina. The ILM, OLM, Bruch's membrane and the sclera are shown as insulators shunted by electrolytic impedances on the left. The result is a series of delay lines between the INM and the ideal signal return point. An additional delay, δ, is encountered between a given location on the retina (within the dashed box) and these delay lines. The black dot indicates a signal probe in contact with the outer surface of the sclera. The return probe is shown connected to an "indifferent location." This probe location makes the test set susceptible to extraneous signal pickup. See text for details.**

The variation in the time constants and the delay associated with the RC networks between the various layers of the retina and the corneal pickup point can cause a number of modeling difficulties. Some of these circuits may play a role in limiting the measured frequency response of the ERG circuit. Cideciyan & Jacobson appear to have introduced their exponential term with a time constant of τ_{rm} into their equation 2 to account for these circuit elements[80]. At the same time, they introduced a fixed time delay given by $\varepsilon = t - \delta$. If these circuit elements impact the frequency response, they necessarily impact the time delay and the shape of the transition from the baseline to the *a*-wave in the composite ERG. The measurement of an LERG, with a penetrating probe, can eliminate many of these circuit elements and produce a waveform more closely resembling those measured by Baylor, et. al. (and others) at the pedicle of a single cell.

16.7.2.1.2 Location of discrete signaling components

Figure 16.7.2-2 (similar to **Figure 3.3.1-3**) provides a more detailed view of the circuitry within the dashed box of the previous figure. In this figure the letters within circles relate to the class of waveform appearing at that point. These classes were individually defined in **Section 11.1.5.1**. The Outer Segment is shown as completely within the protected environment of the IPM. The inner segment of the photoreceptor cell is also shown predominantly within the IPM. The nucleus of this cell is generally shown as located within the INM. It appears that all of the electrical circuitry associated with the photoreceptor cell is located withing the IPM except for the conduit connecting to the pedicel located in the INM. This includes the dendritic structure (usually labeled the microtubules) coursing through the disk stack and entering the Inner Segment through the colax, as well as the distribution amplifier of the cell shown on the IPM side of the OLM. The pedicle synapses with the next neuron of the fundamental signal path (nominally the dendrite, D, of a bipolar cell). This neuron is located completely within the INM. The following ganglion cell may straddle the ILM. However, the intent is to show the axon of this ganglion cell entering the optic nerve by passing through the lamina cribosa, the membrane separating the retina from the optic nerve. Here again it will be assumed that the Activa of the ganglion cell is located within the INM. This signal path is the fundamental path associated with the luminance channel of vision.

[80]Cideciyan A. & Jacobson, S. (1996) Op. Cit.

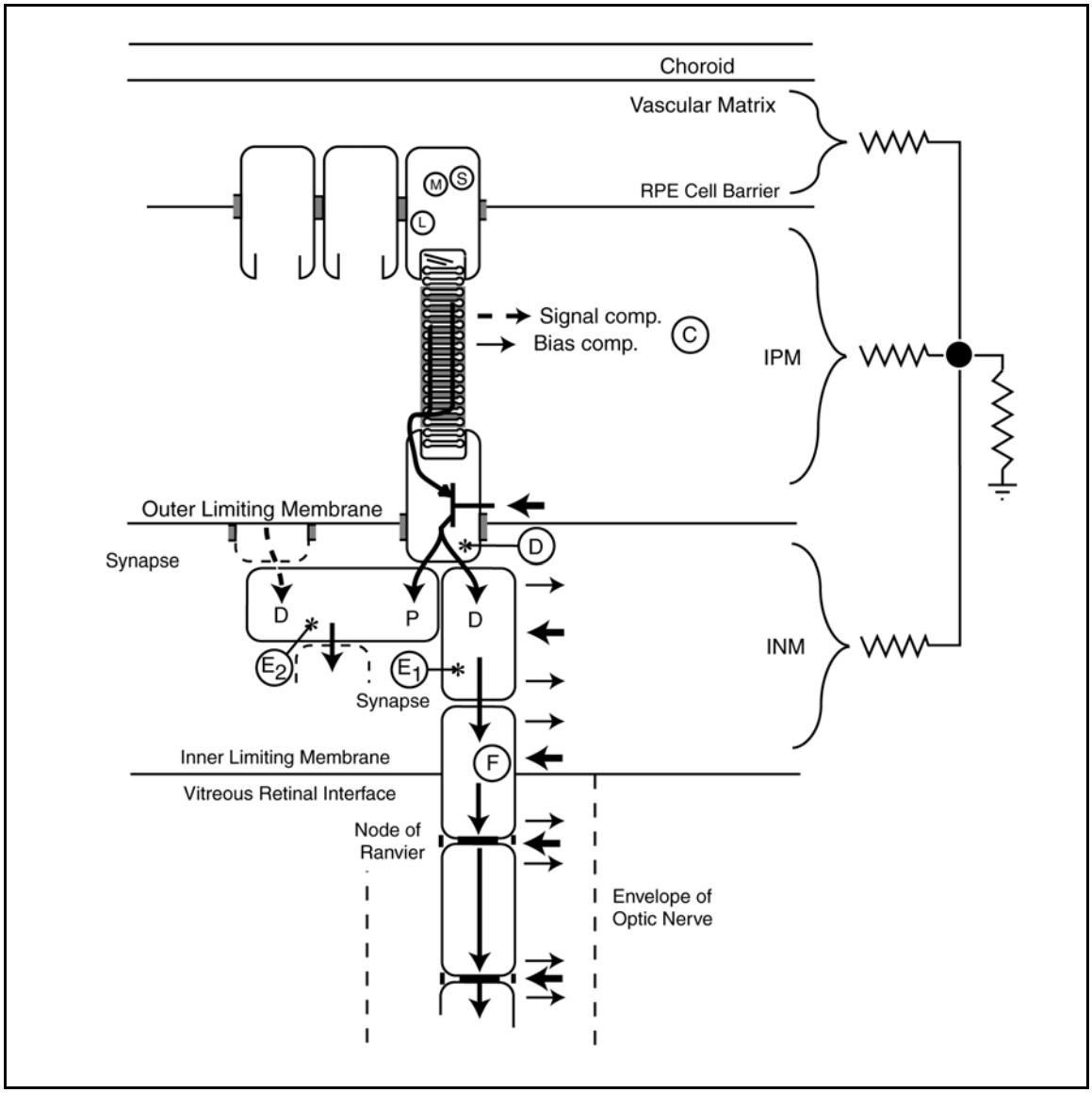


Figure 16.7.2-2 Signal flow within the fundamental signal path of the retina.

The photoreceptor cell may also contribute to one or more chrominance channels by synapsing with one or more lateral cells. A lateral cell is shown to the left of the fundamental path. By accepting signals from two photoreceptor cells of different chromatic types, it begins a chrominance channel as shown here only in outline. Note that the photoreceptor signal is applied to the dendritic, or non-inverting, input to the bipolar cell labeled D. However, in this example, the same signal is applied to the poditic, or inverting, input to the lateral cell labeled P. As a result, two different categories of Class E waveforms must be recognized. The Class E_1 waveform exhibits the same polarity as the Class D waveform. However, the Class E_2 waveform is of opposite polarity to the Class D waveform and its absolute magnitude may be significantly influenced by the input to its dendrite from the other photoreceptor cell shown only in outline.

The series of vertical arrows are meant to describe the signal path of the visual process. In the photoreceptor cell, the Activa of the distribution amplifier is shown explicitly. Its output passes across the synapse to the bipolar lateral cell. An arrow crossing the junction between two cells is meant to imply continuity of current flow. Notice in the case of the ganglion cell, the arrow crosses the synapse at its

dendrite end but does not cross the Node of Ranvier at its axon end. The synapse at the dendrite terminal operates in an analog mode as a low impedance connection, in fact a rectifying connection due to its active properties. However, the Node of Ranvier, employs an Activa in an oscillatory mode or signal regenerative mode. The output current of the Node is independent of , and larger than its input current. The current is shown passing the length of the neural segment within the optic nerve because these axon segments are in fact passive conduits.

The horizontal arrows are meant to designate the flow of current into and out of the individual cells. The thin arrows represent steady currents associated with the electrostenolytic processes powering (and biasing) the various neurons. These occur at specialized surface locations on every dendrite, podite and axon. The heavy horizontal arrows are meant to indicate the areas where non-signal electrons are entering the cells (a conventional current is leaving the cells). The terminology is a bit strained, because the actual electrostenolytic mechanisms are not shown, these arrows could be considered “sources” in conventional electronics terminology. The light solid horizontal arrows illustrate where non-signal electrons are leaving the individual cell. Note that within the IPM and the INM, the source arrows are shown supporting the Activa near the nominal center of each cell. This is not the case within the optic nerve. The axon segments do not contain an Activa. The only active devices are at the Nodes of Ranvier. Thus the arrows are shown in groups of three at these locations. These are the currents that are normally found to be passing through the constricted space of the Node, between the so-called nodal gap and the interneural matrix of the optic nerve.

It is important to note that the only location where the total signal, Class C waveform, passes into the surrounding electrolyte is in the region of the disk stack (the dashed heavy arrow at the start of the signal path). Both signal current and non-signal (bias) current pass into the IPM from the dendritic structure of the Activas of the adaptation amplifiers in this region. The potentials within the various plasma conduits cause currents to flow between each plasma and the surrounding electrolytic matrix. The magnitude of these currents associated with each plasma is relatively small because of the very high impedance of the lemmas of the neurons. However, a local change in the potential of the matrix is caused by these currents. The impedance of the lemma and the matrix at a given point impacts on the nature of the ERG. These impedances have an even greater impact on the LERG, where the probe impedance may be one to several orders of magnitude lower than that of the individual lemma. The signal passing into the IPM has been defined as a Class C waveform. It is fundamentally a composite current waveform. It consists of an amplified copy of the Class A waveform (given by the P/D Equation) at the base terminal of the adaptation amplifier and the dynamic current associated with the adaptation process. The latter current is a result of the vascular/electrostenolytic process at the collector terminal of the Activa. By flowing through the impedance of the IPM, this current generates a voltage that can be measured easily. The signal at the voltage node of the photoreceptor cell, the pedicle, is a Class D waveform. It is a displaced and logarithmically limited replica of the Class C waveform.

As a result of processing by the bipolar and lateral cells, the two categories of Class E waveforms are fundamentally different. The Class E_1 (luminance) waveform is fundamentally monopolar and negative-going. The Class E_2 (chrominance) waveform is fundamentally bipolar. It is positive-going relative to the Class D waveform applied to its poditic input and negative-going with respect to (i. e., of the same polarity as) the Class D waveform applied to its dendritic input. When applied to the inputs of the ganglion cells, these two different categories of Class E waveforms result in two Class F waveforms (action potentials) that have one distinctly different characteristic. The Class F_1 waveform of the fundamental path, generated by a parasol type ganglion cell, is a series of action potentials that is directly related to the amplitude of the Class E_1 waveform, and hence the earlier waveforms as well. The other, Class F_2, waveform generated in the alternate, in this case chrominance channel, is a continuous stream of pulses, even in the absence of a Class F_2 signal. The time interval (and hence the instantaneous frequency) between these action potentials is proportional to the amplitude of the Class E_2 waveform. Laboratory investigations have frequently recorded the Class F_2 waveform. On occasion, the frequency of the action potential stream is observed to go to zero under light stimulation of the associated photoreceptor. While this has frequently been claimed to be a signal that goes off in response to stimulation, it has no direct relationship to luminance. It is describing the chrominance of the stimulant in an entirely different format from that of the luminance channel.

While all of the electrotonic signals associated with a given fundamental path suffer nearly the same delay in traveling radially within the retina. This is not true of the signals traveling through the retina. Each class of waveform appears at a distinctly different time due to the low transport velocity of the plasmas within the various conduits. The individual delays associated with a given Class are relatively constant except for those fundamental paths occurring within the fovea. As shown in many texts, the typical human retina is about 400 microns thick. Except in the area of the fovea, the signal paths from the photoreceptor cells to

the ganglion cells run nominally perpendicular to the plane of the retina. Thus nearly all signal paths within the retina, up to the first axon segment of each ganglion cell, are of equal length. The exceptions are related to the extended length of the conduits (axons) between the distribution amplifier of the foveal photoreceptor cells and their pedicles located outside of the foveal radius within the INM (see Figure xxx, Dowling with overlay). This conduit length can exceed 300 microns whereas the typical photoreceptor axon is less than 100 microns long.

16.7.2.1.3 Complexity associated with test conditions

The time delay associated with the arrival of a series of electrotonic signals from the fovea is significantly different from the time delay associated with a set of non-foveal signals. This difference can be measured by using light stimuli of finite diameter on the retina. However, the use of small light sources leads to low level signals that are frequently noise limited by the test set. A solution to this problem is the use of full field illumination (Ganzfeld illumination). The problem here is that the contribution of the relatively small number of foveal signals is lost in the integration process associated with a single point test probe. This variation in delay can have a deleterious affect on the shape of the composite ERG, particularly in the transition from the baseline to the leading edge of the *a*-wave, but also in the appearance of ripples (sometimes defined as oscillations) within the composite ERG. The above transport delays are not compatible with modeling as a multistage filter.

Finally, investigators frequently introduce a bandpass filter in their test equipment without discussing its phase response versus frequency. The phase response is particularly important in filters with bandwidths spanning multiple octaves. The designation of the number of filter stages in the filter is not sufficient. If the phase response is not a constant with frequency (and most filters are not), additional distortion will be introduced into the recorded waveform[81]. This distortion is typically noted in the area of the toe of the a-wave.

16.7.2.1.4 Other models of signal flow related to the ERG

The vision community has been notoriously lax in the use of comprehensive models. Hood & Birch have provided one of the only models related to the ERG[82]. While showing a familial resemblance to the models of this work, it does not delineate the major sources of delay related to the technique. They also give a brief overview of how the community arrived at its current definitions of the *a*- and *b*-wave of the ERG (their figures 1 & 2). The a- and b-waves were first defined by Granit (1933, 1947). He associated his *a*-wave with the photoreceptors and the *b*-wave with the inner nuclear layer (INL) of the retina. The waves were defined by their appearance in the composite ERG wave. Limited success has been achieved in separating the underlying components of this composite. Fourtes & Hodgkin were on the right track in the 1960's. However, their activity floundered due to the use of inadequate mathematical techniques. The community has not advanced from these techniques to this day. It still relies upon simple algebra instead of complex algebra (the norm when seeking to relate frequency response and time delay in electronic circuits). Lamb entered the picture as part of this team and adopted the position that the *b*-wave started at the negative peak of the *a*-wave. He has pursued this approach to this date. It suffers from a variety of problems related to the time delays associated with the composite ERG. Because of the variable time to peak in the leading edge of the ERG, the beginning of the *b*-wave must also exhibit a variable time delay relative to the excitation stimulus. However, this variable delay must be different than the initial delay associated with the *a*-wave. In fact, the delay between the *b*-wave and the *a*-wave is fixed in time by the transport delay of the neural circuits. It also suffers from the requirement that the *a*-wave, due to an impulse stimuli, be monotonic over an extended period of time (more than 60 milliseconds). This is incompatible with the waveforms, measured by the Baylor & Lamb team in the 1960's, for the response of the photoreceptors to impulses. Hood & Birch continued to adopt the Lamb position. However, their model shows a constant delay (15 ms) related to the INL (part of their P2 generator). Their model does not include any features associated with the delay in the signals from different parts of the retina reaching the corneal location of signal pickup by the test set. Nor does it include any discussion of the potential ground loops associated with the signal, return & ground leads. The “static nonlinearity” of the photoreceptor model of Hood &

---

[81]Storer, J. (1957) Passive Network Synthesis. NY: McGraw-Hill pp 116-121

[82]Hood, D. & Birch, D. (1992) A computational model of the amplitude and implicit time of the *b*-wave of the human ERG. Vis. Neurosci. vol. 8, pp 107-126, fig 8

Birch appears to be the diode at the pedicle of this work. While comparing the number of "filter sections" proposed by Baylor et. al. versus Penn & Hagens in a related paper, they continue to rely upon a multiple stage filter approach[83]. However, they make no effort to define the physical location of such filters.

A more adequate modeling of the composite ERG requires the introduction of complex algebra into the discussion. This modeling must also consider the phase delay within the external filters used in the test equipment. This phase delay and the phase delay associated with the signals traveling to the cornea can exhibit a variation with frequency that normally accounts for the second order departure from the baseline of the measured waveform (the gradual curve instead of a distinct departure associated with an exponential waveform). While desirable, the use of 10 ms flashes in their higher-power test set requires the use of a different mathematical description of their stimulus. It cannot be considered an impulse when it is of the same order of duration as the *a*-wave. There also appears to be a problem with their equation (3). For an exponent of n = 1, defining a single stage filter, this equation gives a value of unity independent of time. The problem is probably a transcription error while contemplating Hodgkin's earlier equation. The exponent should be just n. Hood & Birch do not repeat Hodgkin's caveat that the use of such a simple equation for a multistage filter requires the functional (typically electrical) isolation of each stage of the filter (in Hodgkin's case, by circuits having the isolation capability of a pentode tube).

16.7.2.2 RESERVED

16.7.2.3 Fundamentals and complexities of the ERG

The Electroretinograph is designed to provide detailed electrophysiological data concerning the eye using only external means. It attempts to determine the nominal signaling profile of the visual signals with emphasis on those signals within the ocular globe. The resulting waveforms have been found to vary enormously and to confound logical interpretation in the absence of standardization of the procedures used. The problem of interpretation has been confounded by the lack of a detailed model of what is being measured. By exploiting the information in the previous Section, progress can be made in resolving this problem.

The data collected in Armington is particularly useful[84] for research purposes as well as selected material in Heckenlively & Arden[85]. However, the conclusion drawn throughout these books, not necessarily all by the immediate authors, indicate the difficulties faced in the absence of an adequate model. Three basic problems appear immediately in the analysis of ERG waveforms:

+ Assumptions made in the past about the fundamental waveforms

+ Use of the methods of Kernal Analysis in the absence of fundamental analysis

+ Inadequate attention to the detailed definition of the source of illumination.

16.7.2.3.1 Assumptions concerning fundamental waveforms

A basic problem relates to the assumptions made about the fundamental waveforms aggregated into the composite waveform measured by the ERG. It is absolutely necessary to separate the discussion of ERG composite waveforms obtained under impulse and long pulse illumination conditions. During the 1980's, it became common to think of the initial electrophysiological waveform of the ERG, the so-called a-wave (not the Class A waveform defined here), as being described by a type of hyperbolic function frequently described as a logistic function or Michaelis equation. This was perpetuated by work during that period attempting to describe the a-wave by curve fitting. Similar modeling of the b-wave led to difficulties in combining the two. The subject of the b-wave encroaching on the a-wave is frequently found in the literature. As seen in this work, the fundamental mathematical description of the actual waveforms, combined with the proper combining of their amplified progeny according to the top level schematic of vision, avoids these difficulties. Therefore, this work and this author does not support the recent statement

---

[83]Hood, D. & Birch, D. (1990) The *a*-wave of the human electroretinogram and rod receptor function. Invest. Ophthal & Vis Sci. vol. 31, no. 10, pp 2070-2081
[84]Armington, J. (1974) The electroretinogram. NY: Academic Press.
[85]Heckenlively, J. & Arden, G. (1991) Op. Cit.

by a presumed distant relative[86], that "The hyperbolic function reasonably well describes the relationship of the rod a- and b-wave potentials to stimulus intensity."

For experiments employing long pulse illumination (pulse length longer than 500 ms.) the ERG waveforms can be analyzed based on the response of the eye to two separate changes in illumination, one on-going and one off-going. For long pulse lengths between 50 and 500 ms., the analysis may become extremely complex due to the failure to complete one transient event before the initiation of a second. This situation introduces initial conditions in the differential equations that are difficult to assign.

One conclusion from the following discussion is that the electrolytic environments of inter-neural space can be considered as simple resistive impedances whereas may cell walls must be considered as diodes, of excellent electrical quality. These diodes may or may not be shunted by capacitors and or equivalent resistive impedances associated with the electrostenolytic process.

16.7.2.3.2 Kernal analysis in lieu of rigorous modeling

In lieu of an adequate theoretical model, attempts have been made to mathematically analyze the ERG as the output of a black box describing a fundamentally linear system. Odom provides a good summary of the approach[87]. The method is similar to the chemists use of kinetic analysis to estimate, based at least partially on experience, what is occurring in a reaction. His opening sentence scopes the problem in a manner not accepted in this work. Although the visual system employs a variety of non-linear processes, they are all amenable to precise mathematical description. Only the non-linear operation of the adaptation amplifier has been left to others by this work (in the interest of the efficient use of time, although the author would dearly love to address this problem in the laboratory). Lacking the required model, Odom suggests kernal analysis as a powerful analytical tool. It is such a tool in the absence of more direct methods.

16.7.2.3.3 Importance of illumination characterization

The ERG suffers from a significant dependence on the precise nature of the illumination used to generate it. This is due primarily to the differencing circuits provided by the lateral cells of the retina. Such differencing has not been properly treated in prior works. The difficulty centers on the difference between the Class E_1 and E_2 waveforms defined above due to an individual photoreceptor cell responding to illumination. These waveforms can have virtually identical shape and opposite polarity, in which case they tend to cancel each other. On the other hand, the E_2 waveform can have any value and polarity as determined by a different photoreceptor cell. It can actually have a value of zero. Therefore, to obtain a unique and stable ERG from a given retina, it is necessary to control the experimental parameters that control the relationship between the Class E waveforms. These waveforms are found within the signal manipulation stage of the visual system. Within this stage, the signals have already been stabilized in amplitude excursion by the adaptation amplifiers. Therefore, the control of the parameters impacting the Class E waveforms involves the control of the state of the adaptation amplifiers and the current generated by the P/D process within each chromatic type of photoreceptor. To do this, the total absorbed photon flux (not energy) applied to the various photoreceptors must be controlled. An example will be discussed in the next Section of zeroing the Class E2 waveform in order to obtain a simpler ERG composite waveform.

---

[86]Fulton, A. (1991) Intensity relations and their significance. In Heckenlively & Arden, Op. Cit. Chap. 31, pp. 260-263

[87]Odom, J. (1991) Kernal analysis. *In* Heckenlively & Arden, Op. Cit. Chap. 30, pp. 254-259

The wide range of previously recorded ERG's represent a continuum of test conditions. There are two limiting types of ERG recordings and two illumination options leading to one of those types. In one type, the Class E_2 waveform due to the chromatic lateral cells are positive. In the other type, they are negative. The positive type occurs when the median wavelength of the illumination is near the center of the M-channel absorption spectrum and the standard deviation about that median is small. For the type where the Class E_2 waveform is negative, the median wavelength of the illumination can be as far from the center of the M-channel as possible with the smallest standard deviation as practical (consistent with adequate spectral absorption). This condition can actually be achieved under three scenarios, a blue source (λ_{cutoff} <460 nm.), a red source (λ_{cutoff} > 620 nm.) or a sum of these two source types. If desired, the subject can be chromatically adapted prior to the tests to further optimize the presence of a specific polarity of Class E_2 signals.

The contribution of the Class E_2 waveforms due to the appearance lateral cells is quite small in humans. No attempt has been made in this work to study them or optimize their recording.

16.7.2.3.4 Significance of waveform averaging

As mentioned elsewhere, the use of multiwaveform averaging to overcome the effects of noise or inadequate instrumentation has a negative component. As in the case of dark adaptation studies, averaging frequently obscures details of the underlying process. Such averaging in the case of the ERG, when used with Ganzfeld illumination, completely obscures the difference between the time delay related to the channels related to the foveal photoreceptors relative to that of the non-foveal photoreceptor channels.

16.7.2.3.5 Segmenting the ERG in time

Heckenlively & Arden have included a variety of Chapters discussing different aspects of the ERG and assigning labels to different features commonly found in the recordings over a period of up to five minutes (See figures 9-2, 12.2, 13-1, and 14-1). The papers do not discuss the possible sources of these signals, especially for those occurring after about five s econds. An important finding discussed in the next Section involves the time required for the analog signals generated by a flash illumination, an impulse, to travel from the chromophores of the Outer Segment to the ganglion cells. These signals only require about 100 ms. (300 ms. worst case for the foveal signal paths) to reach the ganglion cells where they are dissipated. The signals recorded by an ERG for times after 100 ms. from the initial signal in the recording are not directly related to signaling of information to the brain, unless there is a long term signal storage capability within the retina that has not been defined in the top level schematic of this work. The existence of such a capability is believed to be unlikely. The author found no sign of such a capability during an extensive, global, review of the literature. It appears that all features occurring in the ERG after 100-200 ms. from the initial feature are related to vascular/electrostenolytic processes related to one of several different fundamental processes. First is the restoration of the quiescent electrical parameters of the retina following the passage of the signaling pulses. The second includes a wide range of non-information signaling related events such as ocular-motor, iris control and shutter control (eyelid related) signals. A possible source may be due to the method of grounding of the retina relative to the rest of the body. The travel time of action potentials between the ganglion cells and the initial stellate cells of the cortex is about XXX. The waveforms, Class G & H, associated with these stellate cells are partial and/or complete reproductions of the Class D & E signals of the retina. Therefore, it is not unreasonable to relate the feature labeled the d-wave, and frequently portrayed as a delayed and somewhat smaller copy of the b-wave (or combined a- and b-wave, see Evers & Gouras[88]) to the reconstructed signals of the stellate cells as recorded by ground currents detected either by the signal electrode of the ERG apparatus near the eye or by the return probe frequently attached to the temple or earlobe. The round trip travel time of such signals could easily reach a second or two, particularly if they were traveling by physical charge transport through the soft tissue of the head, similar to the charge transport velocity calculated below for within the retina.

Based on this analysis, it is predicted that no so-called OP-waves found within 100 ms. of the first feature in an ERG. In this work, these features are associated directly with the collection of artifacts from the action potential generation function of ganglion cells. Their spacing and detailed wave shape leave little doubt that the ganglion cells (probably their dendritic portion because of topology, are the source of these features.

---

[88]Evers, H. & Gouras, P. (1986) XXX Vision Res. vol. 26, pp. 245-254 *also* in Heckenlively & Arden, Op. Cit. figure 12-2

Many of the "features" of the ERG reported in the literature are obtained under less than well controlled conditions, in terms of the requirements developed in this work. The interpretation of these features becomes more difficult the more isolated the laboratory and investigator reporting the feature. A number of the more unusual features given ()-wave labels fall into this category.

Steinberg, Linsenmeier & Griff[89] illustrates another significant aspect of the problem of interpreting the later stages of an ERG recording. They show three recordings of an ERG from a cat taken simultaneously. The upper two traces show the completely opposite polarity of the major features. They also show a nominally horizontal feature in one trace represented by its derivative in the second. The third trace appears to show an intermediate situation consisting of the sum of the above two traces. Lacking a detailed model explaining how both a waveform and its derivative can be summed in a single trace makes comprehensive interpretation of these individual ERG's virtually impossible. Recognizing that the location of the return probe plays nearly as important a role as the signal collection probe is critical to the comparison of ERG's from different laboratories, or even different technicians at different times.

16.7.2.3.6 Decomposing the ERG in frequency

The low frequency components of the ERG are clearly represented by the frequency spectrum of the underlying P/D process.

ERG's are frequently recorded that exhibit high frequency components. These may appear as a fixed frequency ringing or as a series of pulses of varying temporal spacing. The latter have been given the designation oscillatory potentials (OP). The fact that most signals composing the ERG are obtained via capacitive coupling between the plasmas of the neural conduits and the surrounding electrolytes causes these high frequency signals to be emphasized, possibly out of proportion to their importance.

Wachtmeister has provided excellent data related to these phenomena[90]. Most of the traces presented in the literature represent single oscillograph sweeps. This makes it difficult to demonstrate conclusively whether the high frequency components are synchronous with the lower frequency components. The approach of Wachtmeister overcomes this problem by presenting pairs of traces in some situations. Wachtmeister has noted that the presence of the high frequency components is a function of the intensity level of the excitation flash. He used a series of flashes in a complex pattern that requires even closer attention to analyze. He noted that "The energy of the oscillatory potentials remained low when low-intensity conditioning flashes were used. High-intensity conditioning flashes induced a marked gain of energy of the oscillatory potentials." He also noted that in at least some circumstances, "the dominant frequency of the oscillatory potentials decreased evenly from about 150 Hz to 110 Hz as conditioning flashes of moderate and higher intensity were used." In addition, note was made of the variability of the pulse spacing with the first two pulses more spaced than subsequent pairs. Cursory analysis of the data using this work as a background suggests:

+ The impedance environment in which these components are found is entirely unspecified at this time. It appears that the impedance may be changing due to the action of some of the waveform sources. The result could be uncontrolled changes in OP amplitude.

+ For signals below -2.0 log units, the eye was operating in the mezotopic regime or lower as indicated by the change in signal amplitude with illumination level.

+ There are high frequency components that do not appear until after completion of the rise of the waveform associated with the b-wave. Based on his figure 6a, these oscillations appear to be synchronous with the remainder of the ERG. These could be related to the action potentials generated by ganglion cells. They do exhibit the variable spacing expected from the parasol type ganglion cells under these circumstances. The variable spacing suggests that the source is the parasol ganglion cells of the luminance channels. However, the width of these pulses is frequently over 2 ms. This width is not compatible with

---

[89]Steinberg, R. Linsenmeier, R. & Griff, E. (1985) Retinal pigment epithelial cell contributions to the electroretinogram and electrooculogram. *In* Osborne, N & Chader, G. (eds): Progress in Retinal Research, vol r NY: Pergamon Press. *Also* Figure 9-2 of Griff in H & A, pg. 93

[90]Wachtmeister, L. (1972) On the oscillatory potentials of the human electroretinogram in light and dark adaptation. Acta Ophthalmology, vol. 50, pp. 250-269

the typical action potential width of little more than 0.15 ms in humans.

+ There is a waveform that appears to be a constant frequency oscillation occurring in some ERG's prior to the peak of the b-wave. It does not appear after the peak. The oscillations appear to be synchronous with the ERG based on the pair of higher intensity traces in figure 6b. These oscillations disappear at flash intensities below -2.0 log units in his figure 6b. If these characterizations are correct, the OP's are probably due to an instability unrelated to the ganglion cells. Since the ERG integrates the voltage waveforms from millions of individual signaling channels, it must be assumed these oscillations occur generally and are probably a result of a negative impedance associated with the Activas in specific neurons. Since they occur prior to the initiation of the Class D waveform, they are probably an artifact associated with the generation of the Class B & C waveforms by the adaptation amplifiers. This assumption would correlate well with the observation that the high frequency oscillations are a function of the input illuminance intensity. The adaptation amplifier is changing its amplification factor during such intensity changes. The changes in the amplification amplifier result in a change in the negative impedance associated with its transfer function.

The tentative conclusion can be drawn that the high frequency components in ERG recordings are due to two different processes occurring during two different time periods and due to two different causes as noted above. The width of the pulses in human ERG's do not correlate well with the nominal width of action potentials in-vivo, although they do correlate with the expected pulse interval. It is quite possible that the action potentials may be the cause of the oscillatory potentials. The high frequency components of the action potential waveforms are easily applied to nearby undesired circuits through capacitive coupling. Where negative impedances are associated with active devices, this coupling can easily force undesired oscillations, particularly under abnormal operating conditions. They may only occur over a limited range of the operating characteristic of the active device.

16.7.2.4 The nominal ERG composite waveform

By careful design of the ERG recording experiment, much more data appears to be available for research purposes than is normally obtained for clinical purposes. The keys are the careful control of the state of adaptation, the use of flash illumination and the careful choice of the spectral characteristics of the illumination. Location of the return electrode is also important in avoiding non-signal related inputs to the recording. An additional optimization of unquantified value would be the suppression of light reaching the foveal area beyond the small illuminated feature usually used to establish fixation. Avoiding illuminating the fovea may enhance the contrast of the overall ERG, and possibly reduce the need for such extensive signal averaging as found in most experiments.

If the total integrated illumination is kept low, typically in the 10-20 scotopic-td-sec. as commonly described in the literature, the human eye operates in the small signal, linear mode. Under these conditions the composite ERG waveform can be decomposed into the various waveform classes defined in this work with ease, as will be shown below through synthesis.

16.7.2.4.1 Choice of test configuration in experiment design

There are unique opportunities to select an illumination source for an ERG recording that support more precise interpretation. Similarly, there are unique opportunities in the selection of the field of illumination supplied to the eye.

There are three forms of Class E waveforms in the eye, one that is monopolar due to the luminance channel and two that are bipolar due to the chrominance channels. If the bipolar channels could be eliminated, the resulting ERG would be simpler. This can be accomplished by using the New Chromaticity Diagram for Research. The goal is to create a zero net current at the output of the chrominance oriented lateral cells of both the P-and Q- signaling channels. This corresponds to a *perceived* pure white illumination. Such a perception can be obtained in a variety of ways. Under dark adaptation conditions, an illuminant at 7053 K temperature will produce such a situation. This requires the use of a *unique* white light source. Another option is to minimize the radiation in the short wave portion of the spectrum and balance the radiation in the medium and long wave portions so that the output of the lateral cells of the Q-channel is zero.

Although it is customary to use Ganzfeld, or full retina, illumination in ERG recordings, this can be shown to be a non-optimal situation. The object of using a Ganzfeld source is to achieve excitation of as many photoreceptors as possible to obtain signal integration at the point of measurement. This is desirable for two reasons:

+ the combination of the low sensitivity of the recording equipment when attached to the external surface of the eye and the low signal level associated with each individual signaling channel.

+ the obscure fact that the majority of the signals obtained in a contact type ERG are only obtained through capacitive coupling .

If only a single photoreceptor were illuminated for purposes of an ERG, the resulting signal currents would create a signal voltage on the order of millivolts along the signal path but at an impedance level beyond the ability of man to measure from an external point. The ERG actually uses a voltage divider network including a capacitor and resistor in the serial arm of a high pass filter arrangement, similar to figure 2-24 of Armington connected to each signaling channel. (Armington explores conceptually the result of using several different types of simple filters in signaling. The result is useful but not particularly relevant to this situation.) As a result, the measured ERG voltage from the signaling channel associated with a single photoreceptor would be a differentiated copy of the original waveform. The amplitude of this waveform would be in the picovolt range. It is impossible to excite the single photoreceptor to high enough levels to achieve a millivolt, or even a microvolt signal range without destroying the photoreceptor. Remember, the signal has already been stabilized in amplitude by the adaptation amplifier. One solution to the impedance problem, since all of the filter arms containing the capacitors and resistive impedance of the INM are in parallel, is to illuminate a larger number of photoreceptors. This not only raises the signal level, but also effectively eliminates the differentiation due to the effective size of the capacitor (that becomes small relative to the effective size of its associated serial resistor).

> Many investigators have encountered and published figures showing "spikes" at transition points in their ERG waveforms that may be related to the above test instrumentation configuration. Differentiation by a high pass filter is always a challenge in electrophysiological instrumentation.

As will be seen in the next Section, the indiscriminate increase in the illuminated area of the retina has a deleterious side effect. It obscures the difference in the timing of signals obtained from the foveal photoreceptors relative to the extra-foveal photoreceptors. A more optimal solution would be to use a an integrating sphere with a finite diameter hole, or very black area, surrounding a small light used for fixation purposes in the side of the sphere opposite the subjects eye. This "modified Ganzfeld" illuminator would eliminate the data from these channels that was not synchronous with that of the other signaling channels. The result would be an a-focal ERG as opposed to the focal ERG discussed by Birch in clinical terms in H & A.–Chap. 13. In a separate experiment, the signals from just the fovea could be sought. However, this experiment may be more difficult because the signal integration will be limited to their much smaller number.

From a theoretical perspective at this time, the use of background illumination during ERG recording only complicates the interpretation of the results. Background illumination, and any pre-adaptation, should be avoided in experiment planning until better data from highly controlled experiments is obtained.

16.7.2.4.2 Method of waveform analysis

With the model available from this work, it is possible to synthesize an approximation to the measured ERG rather than attempt to disassemble it into its constituents. If the synthesized waveform is adequate, the underlying model is validated. If the waveform is not adequate, a decision can be made as to how to improve the model. The problem initially is to obtain a well characterized ERG. Smith and Lamb have provided a candidate in their Figure 5[91]. It is presented here as the red line in a format similar to, but more compact than that of a horizontal serape as used in Rubin & Walls[92]. Also shown in **Figure 16.7.2-3** are the candidate waveforms by Class taken from this work. When summed, these waveforms provide a composite theoretical ERG which is indistinguishable from the Smith & Lamb sample.

---

[91]Smith, N. & Lamb, T. (1997) The a-wave of the human electroretinogram recorded with a minimally invasive technique. *Vision Res*. vol. 21, pp 2943-2952

[92]Rubin, M. & Walls, G. (1969) Fundamentals of Visual Science. Springfield, IL: Charles Thomas, Publisher. pp 190-196

Smith & Lamb obtained the original ERG using Ganzfeld blue flash illumination under dark adapted conditions. Blue was defined by a Lee #195 filter with a peak in transmission at 450 nm. The caption for their figure was "Rod-isolated a-wave family (NPS)." They chose to give the time relative to the center of the flash command pulse. For this curve, the pulse duration was 333 μsec. Changing the notation to reference the start of the pulse would add 166 μsec. to the delay before the start of the traces. The curve does not show any signal compression at large voltages due to either the distribution or adaptation amplifiers of the photoreceptor cells. This is opposed to their figure 6 that does show such compression. This suggests little signal saturation in the visual system under these conditions.

The original curve showed a delay after the initial blue pulse of 3.52 ms. including the above correction, before any deviation from the baseline. By assuming the waveforms of this theory would completely support the generation of this composite waveform, the task was to determine the amplitude and additional transit delay associated with each of the individual components. The fundamental shape of the Class A waveform was taken from the values shown in the Standard Eye tabulation in Section 16.8 of this work.

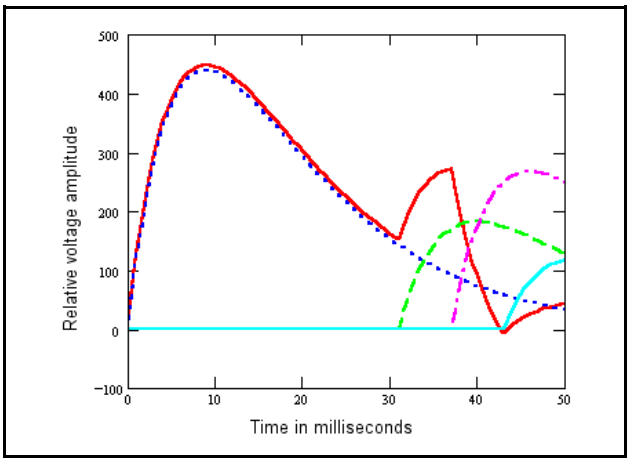


Figure 16.7.2-3 {SCALE} The synthesis of an ERG based on predicted waveforms from this work. [ADD ORIGINAL S & L CURVE and the fith component] The red curve is a slightly smoothed version of the trace marked 250k in Smith and Lamb (1997). The dotted blue trace is a Class C waveform from this work. The green, purple, blue and dashed components are attributed to various waveforms within the signal chain as discussed in the text.

Using the fundamental shape of the Class A waveform to approximate the shape of the subsequent waveforms based on the linearity of the system under *small signal conditions*, **TABLE 16.7.2-1** is obtained. The delays are relative to the first component.

TABLE 16.7.2-1
Individual components of the composite ERG using blue light

Component	Amplitude, rel.	Delay, ms	Proposed Source
1st	+640	0	Class C waveform in IPM
2nd	+267	13	Class D waveform at S-channel pedicel
3rd	–390	16	Class E_2 waveform in chrominance channel, P
4th	+178	18	Class E_1 waveform in luminance channel, R
5th (predicted)	–31.5	19	Class E_2 waveform in appearance channel, Z
6th (predicted)		>22	Class F waveform, action potential

[xxx need to review all of these numbers before publication to be sure the conversion in the "Standard Eye" are consistent.]

To assemble the equivalent composite waveform, the third component must be taken as negative. This is completely compatible with this theory. The Class E_2 waveform associated with the P-channel, a function of the difference M- minus S-channel amplitudes, would be expected to be negative in the absence of a significant M-channel input. The amplitudes shown are relative. All of the individual waveforms assume the original P/D equation was for an absorption coefficient, $\sigma = 0.76$, and a time constant, $\tau = 52.5$ ms. The magnitude of the current due to absorbed flux was given by F= 25 is in photons /sec- micron^2 given by the integrated radiance times absorption as a function of wavelength.

The synthesis may require a 5th component, of negative polarity, starting at about 19 ms. The original data appears to require an additional component to return the composite waveform to near zero volts beyond about 23 ms. However, the data stops near that point. If required, the fifth component would be of the Class E_2 and probably associated with an amercine cell providing spatial signal integration prior to a midget ganglion cell. The human eye is not believed to contain a significant spatial correlation function within the retina. The paucity of such amercine cells would contribute to the low relative voltage amplitude in the integrated signal acquired by the ERG. The introduction of this 5th component would call for an addition to the above figure in the form of an additional lateral cell below the lateral cell shown and receiving its input from the bipolar cell. The graphic arrangement would be identical to the arrangement shown at the pedicel of the photoreceptor cell. This condition was provided for in the original top level schematic of the visual process and is believed to be much more important in predatory animals such as the cats.

> The credence for this level of spatial processing in the retina of the big cats suggests a possible reason for the Zebra's stripes. If the big cats selectively filter their vision , to suppress perception of vertically oriented narrow structures in the scene, and thereby suppress the image information associated with tall grasses to better see animals hidden in the grasses, the Zebra could adopt vertical stripes as an effective defense against big cats.

The signal amplitudes in the above table are all relative because the impedance levels of the individual electrolytic media, the INM and the IPM are unknown, as is the impedance level of the "bridge ground" across which the ERG signal was measured. The fact that these are all relative emphasizes the point that these component voltage levels do not relate to the actual voltages within the neural system. They only reflect the voltage at a given point due to the voltage divider aspects of the network of impedances used to collect the signal.

The delays required to support the synthesis of the composite ERG suggest that:

+ the data was truly acquired under small signal conditions, an initial delay of 3.52 ms. is quite long in the

human and suggests a low illumination pulse based on the delay term of the P/D equation.

+ the distance from the median position of absorption within the Outer Segments to the node of the pedicel within the INM was about 200 microns based on the retinal section of Boycott & Dowling[93]. The calculated charge transport velocity of the analog signal traveling along the unmyelinated axon of the photoreceptor cell would be approximately 15.4 mm/sec. assuming the transit time of the Activa was negligibly small. This is quite slow compared to the transit velocity of the action potential between Nodes of Ranvier but is compatible with the group velocity of action potentials along the optic nerve.

+ the distance between the pedicels of the S-channel photoreceptors and the output of the amercine cells associated with the appearance channel (that are located in the inner lateral layer, a subdivision of the old inner plexiform layer designed to highlight the location of the amercine cells in human retinas (see Boycott & Dowling , above), is about 105 microns based on the same charge transport velocity and the time delay calculated in the above synthesis. This is in excellent agreement with the micrograph.

+ all analog signals related to the original impulse of radiation have reached the ganglion cells and caused the generation of action potentials within about 100 ms. Any subsequent signal recorded by the ERG, such as the features known as b- through m-waves, are not directly related to signaling within the information channels of the eye. They may be signals related to the restoration of the original state of dark adaptation or to other processes substantially unrelated to signaling, such as ocular-motor signals related to maintaining fixation or iris control.

The result of this analysis provides a specification for the signals occurring in the visual path as recorded in this ERG that is considerably more precise than that found by Smith & Lamb using their "delayed gaussian" approximation[94] to the a-wave. The specification defines all of the features of the composite waveform to beyond 50 ms. as opposed to their approximation that only applies to the leading edge of the composite waveform, i. e., during the first six ms. It also shows that the ERG tracks the falling edge as well as the rising edge of each of the waveforms by Class as proposed by this work. This ability to generate an accurate synthetic ERG is strong vindication of the entire electrolytic hypothesis of this work.

Cideciyan & Jacobson[95] presented an alternative to the Lamb & Pugh proposal in 1996. Their analysis surfaced many objections to the Lamb & Pugh analysis based on the cGMP phototransduction cascade hypothesis previously championed by Lamb[96]. Among other points, they noted the failure of the Lamb & Pugh model to fit the data under large signal, or even variable signal, conditions. They provided an alternate model (labeled model 2) based entirely on numerical modeling on a digital computer using Matlab, version 4.2. They still addressed only the leading edge of the total ERG. In contrast to Lamb & Pugh whose model employed "a constant latency and an increasing slope" of the leading edge with higher irradiance levels, these authors proposed a mathematical model with "a constant slope and a shortening latency." In the absence of a detailed theoretical model, they continue to fit curves associated with a higher order differential equation to a first order situation masked by instrumentation noise. The resulting equation of Lamb is a special case of the P/D equation developed in this work ($\sigma \bullet F \bullet \tau = 1.0$) with an additional perturbation term added to compensate for introduction of this equality. Cideciyan & Jacobson introduce additional perturbation terms to further warp their mathematical model to fit the data and claim their solution "fits families of rod-isolated responses better than the delayed Gaussian approximation." Their result calls for a three stage cascaded first order filter and introduces a convolution operation. Such a model cannot accomodate an abrupt departure from a baseline following a pure delay. As shown in this work, neither of these characterizations is appropriate. The theoretical model of this work, and the totally synthesized ERG, shows that the P/D process leads to a solution of the first order differential equation that exhibits both an increasing slope and shortening latency with radiance level. It also accounts for the entire ERG waveform and is not limited to just the rising edge. Neither of these solutions addresses the initial delay adequately. Cideciyan & Johnson assign a majority of it to their test configuration.

---

[93]Boycott, XXX & Dowling, J. (1969) XXX

[94]Lamb, T. & Pugh, E. jr. (1992) A quantitative account of the activation steps involved in phototransduction in amphibian photoreceptors. *J. Physiol*. vol. 449, pp. 719-758

[95]Cideciyan, A. & Jacobson, S. (1996) An alternative phototransduction model for human rod and cone ERG a-waves: normal parameters and variation with age. *Vision Res*. vol. 36, pp. 2609-2621

[96]Lamb, T. (1994) Stochastic simulation of activation in the G-protein cascade of phototransduction. *Biophysical J*. vol. 67, pp 1439-1454

Pepperberg, Birch & Hood[97] published an alternate model, similar to that of Cideciyan & Jacobson in that they propose a four stage cascaded first order filter. They have provided additional new data (mostly from one subject) using Ganzfeld blue irradiation (449 nm peak, 419-466 nm half widths) and artificially dilated pupils. Some tests were performed using a red filter (long wave pass with half amplitude at 605 nm.). No spectral characteristics for the integrated source and filter combinations were given. A background source was operated at a color temperature of only 2650 K. When used at "32 scot-td," it had significant impact on the ERG, reducing their response by 31%. This would suggest, they were operating above the cut-in point of the adaptation amplifier, probably at the higher end of the mezotopic irradiation regime. They developed and advocated a new test protocol using paired flashes. The intent being that "the test flash presented at time zero, is followed at time t by a bright flash that rapidly saturates the rods." It appears that their instrumentation allowed more excitation of the M-channel when using blue light than that of Smith and Lamb. This resulted in poorer definition of the transition between the Class C waveform alone and the subsequent components of the ERG. Also, most of their data was presented at a coarser time resolution. They based their work on the early proposals of Baylor, Hodgkin & Lamb. In that time period, Hodgkin advocated their basic equation which was a special case of that used in this work ($\sigma F\tau = 1.0$) but raised to a higher exponent in order to achieve the desired delay. They conceptually explored a variety of exponents fitting models between the first and 10^{th} order. The goal being to find a solution to the variable latency, variable slope problem presented by the data. The approach of Pepperberg, et. al. doesn't contribute any new theoretical insights.

The synthesized ERG also clearly demonstrates that the nominal definition of the a-wave and b-wave incorporated in the Standard promulgated by the International Society for the Clinical Electrophysiology of Vision (ISCERG) and the National Retinitis Pigmentosa Foundation Fighting Blindness may deserve modification, as least for research purposes beyond the clinic. Under small signal and therefore linear conditions, what might be called the first order kernal, using Odom's terminology, demonstrates that the initial feature of the ERG, usually defined as the a-wave, extends from the baseline to the first peak in the waveform and then back to within 30% of the original baseline before merging with the other components of the composite waveform. This is quite a different profile than that obtained by assuming the a-wave is the nominal height of a logistic function. The second feature, commonly labeled the b-wave begins at this point, after approximately 12.5 ms. in humans at normal 37 C temperature, and extends beyond the relative maximum until approximately 18 ms. During this period, the composite is described by the algebraic sum of the individual Class C, D and E2 components.

The minor deviation between the original ERG above and the synthesized ERG occurs near 12.6 ms. and could be due to the fact that the first component of the actual recorded ERG consists of the Class C waveform of this work, which includes an additional time constant due to the electrostenolytic process. If this time constant were similar to those of these waveforms, it could result in a broadening of the Class C waveform. However, this time constant is much longer than those plotted here, minutes instead of seconds. It is more likely that the notch in the original data may be misdrawn. It would be useful to see the error bars associated with these data points. The waveforms occurring later are true replicas of the Class A waveform and do not contribute any difference between the recorded and synthesized composite waveforms.

There has been frequent reference in the literature suggesting the first peak in the ERG, labeled the a-wave is related to faster cone activity and the second peak, labeled the b-wave is related to slower rod activity. This is clearly not substantiated by the above synthesis which employed no achromatic photoreceptor channel. It shows the a-wave is related to the leading edge of the Class C waveform at the disk stack of the S-channel photoreceptor. This waveform is also known as the early receptor potential. The bulk of the feature known as the b-wave, emanates from either of two sources depending on the test conditions. It may originate at the pedicel of the same photoreceptor, and constitute the Class D waveform, or it may originate at the output of the 1^{st} processing matrix and constitute the Class E_2 waveform.

Because of the technique for recording the ERG, it is possible that some of the features of the composite waveform are due to activity within the cortex, particularly the LGN , the Pretectum or the AOS. The signals returning to the eye from the AOS and other command centers are much closer to the signal lead of the ERG apparatus than are the elements of the retina.

---

[97]Pepperberg, D. Birch, D. & Hood, D. (1997) Photoresponses of human rods *in-vivo* derived from paired-flash electroretinograms. *Visual Neurosci*. vol 14, pp. 73-82

16.7.2.4.3 Observations based on the prototypical ERG

The good agreement between the synthesized and recorded ERG's, especially with regard to the components occurring later in time, suggests that the waveforms were not significantly distorted due to bandwidth limitation in any of the channels prior to the synaptic interface with the ganglion cells.

16.7.2.4.4 Predictions based on the prototypical ERG

With the excellent agreement between the synthesized ERG and the measured ERG of the above figure, an hypothesis can be offered concerning the first appearance of a Class F waveform representing the output of the parasol ganglion cell associated with the fundamental signal path. The principle unknown is the actual location of the Activa within the ganglion cells. The retinal cross-section of Boycott & Dowling clearly shows the location of the nuclei of the ganglion cells just as it does of the photoreceptor cells. However, it does not give any suggestion of the location of the active circuit elements involved. Let the location of all of the Activas within the ganglion cells be located along a line through the centroid of the position of the ganglion nuclei, approximately 100 microns from the INL. Assume also that the charge transport velocity of the path between the bipolar cell and the ganglion cell is the same as in the earlier portions of the same signal path. Under these conditions, the leading edge of the Class D waveform must travel 140 microns at 15.4 mm/sec. It should reach the input terminal of the Activa of the ganglion cell in approximately 9.1 ms. after its formation at the pedicel. The ganglion cell then generates the action potential. The peak of the action potential occurs 0.075 ms. after arrival of the excitation (using the time to switch as equivalent to the time to peak amplitude. The total delay between the start of the Class D waveform and the peak of the first pulse of the Class F waveform would be expected to be 9.2 ms. Under small signal conditions, this would be the earliest expected time for an action potential to be generated. This would correspond to 22 ms. after the beginning of the a-wave in the original figure of Smith & Lamb. Action potentials are typically 0.15 ms. wide for the in-vivo human.

16.7.2.4.5 ERG of other species

A primary consideration in discussing the ERG of other species, and possibly the comparison between ERG's of different species is the effect of temperature. Whereas the quantum processes related to the spectral response of chromophores are only slightly affected by temperature, the electrolytic processes related to signal transmission are highly sensitive to temperature. Both the shape of the Class A waveform due to the P/D process and the charge transport velocity within the conduits of the neural system are strong functions of temperature. At lower temperatures, the very character of the ERG may change through the complete disappearance of the notch between the a-wave and the b-wave.

For comparitive purposes, it is useful to review ERG recordings from other species. Armington[98] includes a figure from Fouchard & Carricaburu (1970) for a variety of flies showing dark adaptation in progress.

[Conclusion from Smith & Lamb curve used in composite waveforms figure is that the individual waveforms are the result of currents flowing through ground bridge impedances which are primarily pure resistances. (not the voltages measured at some nodes.)]

16.7.3 Derivatives of the ERG

Variants of the basic ERG have appeared of two types, those employing invasive procedures and those employing more complex patterns of spatial and chromatic illumination. All of these fall into the category of exploratory at this time because the experiments have not been designed based on a detailed model of the visual system. Improved experiment design offers some promise for the known invasive approaches. There also appear to be some benefit in improving the protocols used in the so-called focal, and afocal (defined in the preceding section) ERG's. It will require careful experiment analysis and design to determine whether the more complex pattern approaches provide meaningful data. In general, all of these design activities should focus on the fully dark-adapted retina until a firm experimental foundation is achieved.

16.7.3.1 Invasive techniques, the Local Electroretinographs (LERG's)

Based on the above discussion, it is natural for the experimenter to want to obtain more precise ERG records and to trace individual features in those recording to their source. The logical next step is to invade the subject in order to obtain a local ERG. If planned properly, improved results can be obtained. However, there are several tradeoffs. Whether the signal probe remains in inter-neural space or penetrates a specific cell makes a major difference. In either case, the one advantage of signal averaging over a large ensemble is lost. The signal voltage may be larger than that from a regular ERG. However, it may also be much noisier. In addition, it may be extremely difficult to determine where the probe is in the topology of signal space. Man made probes are typically an order of magnitude larger and contribute large amounts of capacitance to any electrical contact made to a cell (or multiple cells simultaneously). To locate the cell contacted by the probe may require dye injection and later sacrifice of the animal to determine where the probe was located morphologically. Currently, comparing its morphological location to the morphological data base says virtually nothing about its location in signal space. This work is specifically designed to increase the correlation between our knowledge of the morphological and signal spaces. The figure referenced above by Steinberg, et. al. is particularly enlightening concerning the problems to be encountered when the probe is in inter-neural space. The signal obtained may consist of contributions from many signal sources located in relatively isolated signal spaces. The resulting impedances between the various sources may not be as expected. The resulting summation may involve both signals and their derivatives simultaneously.

It should be noted that the use of a coaxial probe is not always a solution to the location of the signal return lead. In a number of cases in the literature, the probe tip appears to be in electrical contact with the electrolyte of one cell while the return lead is in contact with the electrolyte of an adjacent cell instead of the anticipated inter-neural matrix. Interpretation of the resulting data can be perplexing.

---

[98] Armington, J. (1974) Op. Cit. pg. 119

Armington discusses a variety of these operational considerations[99].

Yang, et. al. have provided LERG's from the goldfish that show the generator waveforms due to individual spectral channels and how they merge as a function of time between the spectral flashes[100]. The paper also provides dark adaptation data.

See probe data in retina folder (depth profiles)

16.7.3.2 Non-invasive techniques, variations in spatial illumination

Section 16.7.2 introduced the subject of optimizing the illumination pattern to improve on the basic Ganzfeld illumination approach. Besides the approach labeled the "modified Ganzfeld" where the fovea was not illuminated in order to improve the clarity of the common ERG, the alternative approach is also available. The foveal ERG illuminates only the fovea in order to study only the foveal signal channels in greater detail. This separation into the modified, or afocal, and the foveal approaches can be justified based on the facts illustrated in the top level schematic and the cross-section of the human retina. The axon conduits in the photoreceptors of each group are of different length and the signal paths go to different cortex locations. In the case of the foveal ERG, it may improve the experiment if the dimension assumed for the fovea is minimized, even to the extent of only illuminating the foveola.

16.7.3.3 Non-invasive techniques, variations in chromatic illumination

This category addresses the use of different chromatic illumination for both the test probe and also for the pre-adaptation. It could also address the subject of backgrounds after pre-adaptation but data from the first two approaches must be well understood before proceeding to this level of sophistication. By careful planning, the performance of individual chromatic photoreceptor channels can be significantly impacted through pre-adaptation. This can significantly impact the signal levels of both the Class D and Class E waveforms and therefore affect the shape of the composite ERG obtained. In general, it will not affect the Class C waveforms since they are generated before the adaptation process. Care must be taken that the test probe illumination remains low to avoid saturation in the adaptation amplifier but more importantly the distribution amplifier of the photoreceptor cells.

16.7.3.4 Non-invasive techniques, variations in spatial patterns of illumination

Many experiments employing patterns have been performed psychophysically to explore the spatial frequency response of the visual system. Because of limited precision in experiment design, the results from these experiments have been less than conclusive and consistent. The approach has been expanded into the electrophysical arena with similar results. In more careful design, particularly with regard to the actual spectral content of the illumination patterns, better results can be expected. The ultimate design would select the spectral characteristics of the illumination patterns to cause either pure chromatic change with no luminance change in the visual pathways or the opposite situation. Both promise to provide less ambiguous information than merely using poorly defined "red" and "green" lights.

Most of the experiments in this category will use checkerboard patterns and test for the performance of the signaling channel with regard to signal transmission and the cortex for signal perception. Because of the limited presence of amercine cells associated with appearance channels in the human, experiments attempting to determine the capability of the human retina to encode pattern information have seen little success. This is probably inevitable.

16.7.4 Analyses of the ERG

Analyses of individual ERG waveforms requires a detailed model of the visual system in order to interpret the observed features. Lacking such a model, the literature provides a confusion of claims and a variety of infrequently observed features. Careful reading is often required to ascertain the precise test conditions used by a specific researcher. As an example, Hagins used a method of normalization in his caricature that

---

[99]Armington, J. (1974) Op. Cit. pg. 170

[100]Yang, X-L., Tauchi, M. & Kaneko, A. (1983) Convergence of signals from red-sensitive and green-sensitive cones onto L-type external horizontal cells of the goldfish retina. *Vision Res*. vol. 23, no. 4, pp. 371-380

obscures the very nature of the original signals[101]. Frequently, analysts do not differentiate between data collected based on an impulse stimulus from that collected using a step stimulus.

16.7.4.1 ERG/LERG Signals not defined above

16.7.4.1.1 Definition of the various -waves used clinically

Several attempts have been made to decompose the ERG into its components without relying upon a physical model of the underlying mechanisms. Granit originally defined a series of π(n)-waves[102]. These had little relationship to any underlying mechanisms. Subsequently, these waves were defined using p or P instead of π and either Arabic or Roman numerals for n.

Brown, et. al. appear to have originally defined the features of the ERG that were subsequently adopted for use in the clinic[103]. Unfortunately, their work was clearly exploratory. Their definition of the *a*-wave, as corresponding to only the rising portion of the initial ERG response has led to many subsequent misinterpretations. It is only the leading edge of the Class C waveform defined here. The *a*-wave resulting from an impulse stimulus cannot be properly interpreted based on a standalone Michaelis equation. Similarly, the other conventionally defined "waves," such as the *b*-wave, do not correspond to physically identifiable portions of the ERG. The various waves are defined in detail in **Section 11.1.5.2**.

Hood & Birch re-analyzed the definition of the above features of the ERG in a series of papers[104]. They note, "But, according to the analysis of Granit and others, there is no *b*-wave generating mechanism.." Their interpretation of the Granit model assumed an initial response generated in the "receptors" and described by a Michaelis equation. The following response was presumed to be generated in the INL. Their emphasis was on how the commonly defined *b*-wave underestimated the amplitude of the underlying mechanisms. Figure 1 of the 1992 paper shows a caricature of an ERG based on an impulse response. However, their figure 2 of a simplified Granit Model appears to be based on a step stimulus model.

Hood & Birch adopted a mathematical model that summed signals from the IPM and the INL that were fundamentally different in shape. They did not introduce any time delay to account for the finite transit velocity of the electrolytic currents of the neural system. They noted that, "Unlike the rod receptor, retinal physiology has yet to provide a computational model of inner nuclear layer activity. The nature and characteristics of each component of the model were chosen based largely upon what seemed to be needed to fit the b-wave and P2 data analysis." This work provides the necessary physical and mathematical model.

16.7.4.1.2 The putative Early Receptor Potential

One feature of the ERG and LERG has not been discussed above. An Early Receptor Potential (ERP) is frequently observed in actual measured waveforms, and frequently included in caricatures, of ERG's. This feature is frequently bipolar and exhibits a total duration measured in fractions of a millisecond. Armington reports (pg 120) Smith & Brown measured ERP's with durations of at most a few microseconds in recordings from *Limulus*. Such short duration pulses are difficult to reconcile with a network that has a bandwidth on the order of 100 Hz or less. The graphic provided does not illustrate such a short duration pulse. Instead, it appears to shows one Class D waveform (with a latency of about 33 ms) and one Class E waveform (with a delay of an additional 25 ms). Th data would suggest a combination of retinula and eccentric cell waveforms. They say intense stimuli were required to observe this waveform. The ERP can

---

[101]Hagins, W. (1965) Electrical signs of information flow in photoreceptors. *In* Sensory Receptors. Long Island, NY: Cold Spring Harbor Laboratory of Quantitative Biology, vol. 30, pp 403-418

[102]Granit, R. (1933) The components of the retinal action potential in mammals and their relation to the dischared in the optic nerve. J. Physiol. vol. 77, pp 207-239

[103]Brown, K. Watanabe, K. & Murakami, M. (1965) The early and late receptor potentials of monkey cones and rods. *In* Sensory Receptors. Long Island, NY: Cold Spring Harbor Laboratory of Quantitative Biology, vol. 30, pp 457-482

[104]Hood, D. & Birch, D. (1992) A computational model of the amplitude and implicit time of the *b*-wave of the human ERG. *Visual Neurosci*. vol. 8, pp 107-126

often be associated with ground currents in the test set, particularly when a flash tube is used as a source of stimulation. However, its ubiquity suggests an alternate source may be present. As shown in Armington (pg 249), the ERP only occurs at the start of a long stimulating pulse. This suggests it is not associated with the on and off transients associated with the stimulus. However, this does not rule out the test set as a source. Flashtubes require a driving pulse at onset but not at termination. Termination is usually determined by the flash tube circuitry. Heckenlively briefly discusses the ERP in a clinical context[105]. He describes it as occurring in the first two milliseconds in humans after intense stimulation. He does not provide any waveforms. He appears to be describing the normal Class C waveform when it is well separated in time from the later waveforms. This occurs normally when the latency of the Class C waveform is less than the delay between the Class C and later waveforms.

Three early papers in volume 30 of the Cold Spring Harbor Symposia discuss the early investigations relative to the ERP[106]. The investigators used a wide range of techniques and conditions. Brown, et. al. label the very fast initial pulse in their waveforms an artifact. Page 494 by Pak shows the temperature dependence of the so-called ERP is predicted in detail by the P/D Equation of this work. This would suggest the ERP in this investigation was in fact the typical Class C or D waveform (depending on probe location) of any animal.

16.7.4.1.3 The putative oscillations in the a-wave

A second feature not discussed above is an oscillation superimposed on an ERG/LERG. No discrete source of these could be defined. However, it is not always clear these features are oscillations. They sometimes appear as plateaus associated with the leading edge of the ERG/LERG. When in this form, they can be easily associated with the variation in delay associated with signals collected from different areas of the retina. Rubin & Walls provide one paragraph on this subject but no definitive waveforms[107]. Wachtmeister has reported more broadly on these oscillatory potential (OP) and associates them with the b-wave of the conventional ERG[108]. He notes the difficulty of observing these responses in unfiltered ERGs as their threshold is 2-3 log units higher than that of the b-wave. He suggests a bandpass filter of 80-500 Hz be used to accentuate the response. He also indicates they are difficult to observe unless stimuli with an abrupt onset are used. These parameters suggest the OP's are frequently an artifact of the filtering. Later, he observes no prominent OP's can be recorded in response to a single flash in the dark adapted eye. "Special adaptational conditions are a prerequisite to maximize the oscillations."

16.7.4.2 Other aspects of ERG/LERG analyses

The model provided here calls for much more orderly data collection if a complete understanding of an ERG is to be obtained. Since the Class E waveforms can reverse in polarity as a function of the illumination used as a stimulus, it is strongly recommended that ERGs be collected using monochromatic light, preferably near 532 nm. To minimize the smearing of the data in time, the illumination stimulus should be limited to the diameter of the fovea, or otherwise tailored to avoid the variations in time delay associated with the merging of foveal and non-foveal data.

As indicated earlier, the inter photoreceptor matrix and the inter neural matrix are separated by the outer limiting membrane, and are both electrolytes of limited conductivity. They also exhibit limited charge transport velocity. Because of this, there is a voltage difference between each of the local grounds and also between any of the local grounds and the Reference ground, where ever it is. This voltage difference and the associated time delays can cause unrecognized perturbations, even oscillations, in the data collected. It is very similar to the problem encountered in high frequency radio work where the inductance of the metallic ground plane, a simple flat sheet of metal, must be considered in both the layout of the components and in the measurements made.

16.7.4.2.1 Frequent difference between ERG and LERG waveforms

The conventional ERG is an analog plot of a conglomeration of waveforms that are summed under largely

---

[105]Heckenlively, J. (1988) Retinitis Pigmentosa. NY: J. B. Lippincott, pg 10
[106]---- (1965) Long Island, NY: Cold Spring Harbor Laboratory of Quantitative Biology, vol. 30, pp 410-500
[107]Rubin, M. & Walls,. G. (xxx) Fundamentals of Visual Science. Springfield, IL: Charles C. Thomas, pp 210-211
[108]Wachtmeister, L. (1991) Oscillatory potential recording. In Heckenlively J. & Arden, G. Op. Cit. Chap 41

uncontrolled but repeatable conditions. This has made the detailed understanding of the ERG extremely difficult. Nevertheless, it has been found useful over the years and is widely used. Although the so-called a-wave has come to be associated with the signals emanating ***en-mass*** from the photoreceptor cells, the end of the a-wave under impulse conditions has not been recognized in the clinic, except as the ERP under very high stimulus conditions.

In the ERG, the initial waveform is due to the Class C waveform generated in the IPM of the retina. In the case of the LERG, the measured waveforms tend to be simpler, since the probe only interrogates a small area of the electrolytic matrix of the INL. The initial waveform tends to be the Class D waveform generated in the INL of the retina. This is because the Class C waveform is generated on the other side of the OLM acting as an insulator. This difference needs to be observed in future research.

16.7.4.2.2 The location of the indifferent electrode is critical

The above paragraphs illustrate a problem when making ERG and LERG recordings. It is normal to say that one electrode was attached to an "indifferent" location on the front of the eye, the side of the eye, the back of the sclera behind the retina, the forehead, on the temple, or on the ear lobe depending on whether the experiments are *in-vivo* or *in-vitro*. For the typical *in-vivo* ERG[109], one electrode is attached to the outside of the eye by a contact lens, the second electrode is connected to an "indifferent" location on the forehead and a third "ground" electrode is attached to the ear lobe. The typical measurement is then the voltage difference between the outside of the front of the eyeball and the forehead. This difference is dominated by the net current flowing down the optic nerve in parallel with any net current flowing through the eye muscles. It is remarkable that any useful information is acquired from the illumination of a small spot of the retina under these conditions; there are dozens of current loops operating within the eye that are neither suppressed or measured individually by this technique. Obviously, the ERG has been found useful in clinical tests. In research, the interpretation of the ERG waveforms has been the subject of considerable discussion over the years; with many theoreticians explaining the waveforms based on amplitude discrimination. However, the circuit diagram of the "straight through" signal path allows the ERG to be interpreted with considerable precision using time discrimination--even though the sum of a large number of signal pathways must be averaged.

Based on the complete P/D equation presented in this model, it is now possible to factor out the entire waveform related to the output of the PC and present the remaining component of the b-wave. This component can then be studied in greater detail.

16.7.4.2.3 The underlying features of the ERG

It is important to differentiate between ERGs based on impulse and long pulse stimulation. While the leading edge of the measured data remains very similar in both cases, the remaining portion can vary significantly.

Granit was partly right, the initial part of the ERG originates in the photoreceptors and the second part originates in the INL. However, parts of both of these signals originate in the photoreceptors. The initial part of the ERG, the Class C waveform, is obtained from the dendrites of the photoreceptor cells located in the IPM. The second part of the ERG, the Class D waveform, is obtained from the pedicles of the same photoreceptors located in the INM (INL). The Class D waveform is a delayed and slightly modified (depending on signal amplitude) form of the Class C waveform. It is delayed due to the low transit velocity of the electrotonic signal along the axon of the photoreceptor cells. The Class E waveforms are again delayed and modified versions of the Class D waveform. Here, the modification can be more complex when performed in the lateral cells of the INM.

16.8 Nominal values of constants in human vision

As part of this work, a wide variety of parameters associated with the human visual system have been

---

[109]Cideciyan, A. & Jacobson, S. (1996) An alternative phototransduction model for human rod and cone ERG a-waves. Vision Res. vol. 36, pp. 2609-2621

considered and or computed. All of these parameters have been accumulated and placed in **Appendix L: The Standard Eye.**

TABLE OF CONTENTS 4/30/17

16 The Equations of Vision and their confirmation 2
16.1 Introduction 2
16.1.1 Background 3
16.1.1.1 Terminology 3
16.1.1.1.1 An equation versus a function 3
16.1.1.1.2 Illuminance versus irradiance 3
16.1.1.1.3 Large signal versus small signal conditions 4
16.1.1.1.4 Additive versus subtractive color mixing 4
16.1.1.1.5 Photometry versus radiometry 4
16.1.1.1.6 Luminous efficiency and “Quantum efficiency” 4
16.1.1.2 Glossary 5
16.1.1.2.1 Terms independent of spectral composition 5
16.1.2 Graphical presentations 101 6
16.1.2.1 The problem of a color pigment equilateral triangle 7
16.1.2.2 Extension of the color triangle to an equilateral tetrahedron 7
16.1.2.3 The problem of the C.I.E (1931) Chromaticity Diagram (as an example) 7
16.1.2.4 Introduction of a conformal color space 8
16.1.2.4.1 Basis for conformal transformations used in chromaticity diagrams 9
16.1.2.5 The consequence of changing from a wavelength to a frequency baseline 10
16.1.3 Introduction of feedback theory 10
16.1.3.1 The requirement for Z--plane analysis when stage 3 circuits are involved 11
16.2 Fundamental block and schematic diagrams of vision 11
16.2.1 The architecture of the visual system 12
16.2.1.1 The Stage 1/Stage 2 interface 12
16.2.1.2 Table of channel designations 13
16.2.1.3 Location of major nonlinearities, asymmetries and gain variations . . 14
16.2.1.3.1 Nonlinearities in vision 16
16.2.1.3.2 Asymmetries in vision 17
16.2.1.3.3 Variations in amplifier gain in vision 17
16.2.1.4 Table of coefficients 17
16.2.1.5 Spectral parameters of the chromatic channels 18
16.2.1.5.1 Limitations on difference spectra 18
16.2.1.6 The circuit architecture of the photoreceptor cell. 19
16.2.1.7 The formation of the saliency map 21
16.2.2 Old and new equations of vision 21
16.2.2.1 Previous obsolete conceptual equations 21
16.2.2.2 Equivalent steady state functional equations 22
16.2.2.2.1 The equations of stage 1, the signal detection stage 22
16.2.2.2.2 The equations of adaptation within stage 2 23
16.2.2.3 The temporally simplest functional equations 23
16.2.3 The equations at the pedicles of the photoreceptor cells 24
16.2.3.1 The output currents at the pedicles 24
16.2.3.2 The output voltages at the pedicles 24
16.2.4 The equations of stage 2, signal processing 25
16.2.4.1 The signals at the S-plane (perceptual plane) of vision 25
16.2.4.2 The transfer of signals between the pedicles and the S-plane 25
16.2.4.3 The signals of the S-plane in matrix form 27
16.2.3.4.1 A simplified (degenerate) matrix 27
16.3 The functional equations of each signaling channel 28
16.3.1 Practical and experimental limitations 29
16.3.1.1 Subdividing the model 29
16.3.1.2 Eliminating marginal channels 29
16.3.1.3 Quantifying the laboratory stimulus 29
16.3.2 The luminance equation of the tetrachromatic and aphakic eye 31

16.3.2.1 The luminance equation within the photopic region 32
16.3.2.2 The spectral sensitivity of the aphakic human eye 32
16.3.3 The luminance equation for the blocked tetrachromat (including humans) 33
16.3.3.1 Losses in the physical optics of the eye 33
16.3.3.1.1 Constituents of transmission loss 36
16.3.3.1.2 Equation fitting Said & Weale data 37
16.3.3.1.3 Observed variations from the Said & Weale data in humans 37
16.3.3.1.4 Macular absorption by the field lens 38
16.3.3.2 The spectral sensitivity of the normal human 39
16.3.4 The Chrominance Equations of vision 40
16.3.4.1 Experimental considerations 40
16.3.4.2 The general form of the chrominance equations 40
16.3.4.3 The chrominance equations in the photopic region 41
16.3.4.4 Color constancy within the photopic region 41
16.3.4.5 Detailed equations of the chrominance channels 41
16.3.4.5.1 Cross coupling between P– and Q–channel signals 42
16.3.4.5.2 Color abnormalities due to neuron bias errors 43
16.3.4.5.3 Loss of color perception at reduced stimulus intensity . . . 43
16.3.4.5.4 Creating a color continuum from the P– & Q–channels . . 45
16.3.4.5.5 An anomaly confirming the tetrachromatic concept 45
16.3.5 The appearance equations of vision 46
16.3.5.1 Spatial signal processing related to the extra-foveal retina 47
16.3.5.1.1 The mechanism of spatial processing in the retina 47
16.3.5.2 Spatial signal processing related to the foveola 47
16.3.5.2.1 The human appearance channel associated with the foveola 47
16.4 The temporal equations of vision 48
16.4.1 Temporal equation of signal generation in stage 1 49
16.4.1.1 The Photoexcitation/De-excitation, P/D, Equation 49
16.4.1.1.1 The scaling factor 51
16.4.1.1.2 The delay term 52
16.4.1.1.3 The amplitude response of the P/D Equation 55
16.4.1.2 Evaluating the total P/D Equation under impulse stimulation 58
16.4.1.2.1 Evaluating the P/D Equation 58
16.4.1.2.2 Comparing the total P/D Equation to measured impulse data 59
16.4.1.2.2 The potential measured within the IPM–Class C waveform 61
16.4.1.2.3 The potential measured at the pedicle–Class D waveform 62
16.4.1.3 Evaluating the total P/D Equation under finite pulse stimulation . . . 62
16.4.1.4 Evaluating the total P/D Equation with a background present 62
16.4.1.5 An extended P/D Equation 63
16.4.2 Temporal equation of signal adaptation in stage 1 63
16.4.2.1 The equations of adaptation 65
16.4.2.1.1 The conceptual equation of adaptation EMPTY 65
16.4.2.1.2 The electro-vascular characteristics of adaptation 65
16.4.2.1.3 The operation of the adaptation amplifier 66
16.4.2.1.4 The equations of the adaptation amplifier–the Exposine function 67
16.4.2.1.5 Dark adaptation characteristic computation worksheet . . 70
16.4.2.2 Comparison of adaptation with the data base of R. E. Records 72
16.4.2.2.1 The nominal characteristics for dark adaptation EMPTY 72
16.4.2.2.2 Evaluating the parameters in dark adaptation EDIT 72
16.4.2.2.3 The nominal characteristics for light adaptation 76
16.4.2.2.4 Evaluating the parameters of light adaptation 76
16.4.3 The generator potential at the PC pedicle EMPTY 77
16.4.4 Temporal equation of signal manipulation in stage 2 EMPTY 77
16.4.5 Temporal equation of signal projection in stage 3 77
16.4.5.1.1 Nominal waveforms of signal manipulation 82
16.4.5.1.2 The axon as a transmission line 84

16.4.5.1.3 The ganglion cell circuit- - (time div. modulation in two forms) EMPTY 85
16.4.5.1.4 Signaling parameters of the projection stage EMPTY . . . 85
16.4.6 Oculomotor equations of vision 85
16.4.7 Important postulates of this theory 85
16.4.7.1 Location of the plane of perception 85
16.4.7.2 Equality of signal amplitudes at pedicles 85
16.4.8 Extensions to the fundamental equations 86
16.4.8.1 The variation in performance due to the physiological optics 86
16.4.8.2 The variation in performance due to the adaptation amplifiers 86
16.5 A summary of signaling in the visual system 86
16.5.1 Collected impulse waveforms of vision 88
16.5.1.1 Waveforms up to the stellate cells of the midbrain 89
16.5.1.2 Waveforms found within the commissure and extraction engines . . . 90
16.5.2 Cumulative time delays in vision 90
16.5.2.1 Time delays up to the midbrain 90
16.5.2.2 Time delays in the alarm channels 90
16.5.2.3 Time delays in the cognitive channels 91
16.5.3 Example of cumulative time delay 91
16.5.3.1 A cumulative time delay for luminance 91
16.5.3.1.1 Variations from the baseline 93
16.5.3.2 A cumulative time delay for chrominance 94
16.6 Evaluation and Interpretation via Psychophysiological Experiments 94
16.6.1 Intermediate complexity situations 95
16.6.1.1 The Complete Luminosity Function 95
16.6.1.2 The Dark Adaptation Function 96
16.6.2 Complex situations 96
16.6.2.1 Evaluation of flicker tests 96
16.6.2.1.1 Flicker testing with broadband illumination 96
16.6.3 Evaluation using Fourier Transform spaces 97
16.6.3.1 The general Fourier Transform 97
16.6.3.1.1 The one-dimensional Fourier Transform 97
16.6.3.1.2 The two-dimensional Fourier Transform 98
16.6.3.1.3 The optical solution to the Fourier Transform 98
16.6.3.2 Parameters related to Moire, sampling and correlation 99
16.6.3.3 Sampling at the basic level 100
16.6.3.4 Sampling and image reconstruction 101
16.6.3.5 Application of Fourier and Moire concepts in vision 104
16.6.3.5.1 The Fourier frequency space for crossed fringe 104
16.6.3.5.2 The Fourier frequency space for crossed fringes applied to a nonlinear mechanism 106
16.6.3.5.3 The Fourier frequency space for the foveola 106
16.6.3.3 Measurements comparing M– and L–channel contrasts 108
16.7 Evaluation and Interpretation via electrophysiological Experiments 109
16.7.1 Evaluation and Interpretation of Ganglion Cell Output 109
16.7.1.1 The scenario 110
16.7.1.3 The data and its key features 111
16.7.2 Evaluation and Interpretation of ERG waveforms 114
16.7.2.1 A comprehensive signal flow, signal delay and grounding plan . . . 115
16.7.2.1.1 The physical arrangement of circuit elements in the retina 115
16.7.2.1.2 Location of discrete signaling components 117
16.7.2.1.3 Complexity associated with test conditions 120
16.7.2.1.4 Other models of signal flow related to the ERG 120
16.7.2.2 RESERVED 121
16.7.2.3 Fundamentals and complexities of the ERG 121
16.7.2.3.1 Assumptions concerning fundamental waveforms 121
16.7.2.3.2 Kernal analysis in lieu of rigorous modeling 122
16.7.2.3.3 Importance of illumination characterization 122
16.7.2.3.4 Significance of waveform averaging 123

16.7.2.3.5 Segmenting the ERG in time 123
16.7.2.3.6 Decomposing the ERG in frequency 124
16.7.2.4 The nominal ERG composite waveform 125
16.7.2.4.1 Choice of test configuration in experiment design 125
16.7.2.4.2 Method of waveform analysis 126
16.7.2.4.3 Observations based on the prototypical ERG 131
16.7.2.4.4 Predictions based on the prototypical ERG 131
16.7.2.4.5 ERG of other species 132
16.7.3 Derivatives of the ERG 132
16.7.3.1 Invasive techniques, the Local Electroretinographs (LERG's) 132
16.7.3.2 Non-invasive techniques, variations in spatial illumination 133
16.7.3.3 Non-invasive techniques, variations in chromatic illumination 133
16.7.3.4 Non-invasive techniques, variations in spatial patterns of illumination 133
16.7.4 Analyses of the ERG 133
16.7.4.1 ERG/LERG Signals not defined above 134
16.7.4.1.1 Definition of the various -waves used clinically 134
16.7.4.1.2 The putative Early Receptor Potential 134
16.7.4.1.3 The putative oscillations in the a-wave 135
16.7.4.2 Other aspects of ERG/LERG analyses 135
16.7.4.2.1 Frequent difference between ERG and LERG waveforms 135
16.7.4.2.2 The location of the indifferent electrode 136
16.7.4.2.3 The underlying features of the ERG 136
16.8 Nominal values of constants in human vision 136

LIST OF EQUATIONS

The indexing program does not treat Greek letters properly. Do not rely on the following equations. See the equation in the text. Lower case l has replace lambda in the following equations. [xxx eliminate from PDF version.]

$\mathbf{Z}(\,) = a_0(\,) + ga_1(\,) + ha_2(\,) + ia_3(\,) + ja_4(\,) + \ldots$ Eq. 16.2.1-1 14

$Q = xR + yG + zB$ Eq. 16.2.2-1 21

$r^x_k = \S\, R_k(\,l_n)E(l_n\,)\,S^x(l_n\,)$ Eq. 16.2.2-2 22

$r^x_k = \S\, E(l_n\,)\,S^x(l_n\,)O_k(l_n\,)\,R_k(\,l_n)\,Dl$ Eq. 16.2.2-3 22

$r^x_k = \S$ Integral $E(l_n\,)\,S^x(l_n\,)O_k(\,l_n)\,R_k(\,l_n)\,dl$ Eq. 16.2.2-4 22

$s^x_k = Kr^x_k = \S$ Integral $E(l_n\,)\,S^x(l_n\,)O_k(\,l_n)\,R_k(\,l_n)\,dl$ Eq. 16.2.2-5 23

$s^x_L = Kr^x_L = [\text{Integral } E(l_n\,)\,S^x(l_n\,)O(l_n\,)\,R_L(\,l_n)\,dl]^2$ Eq. 16.2.2-6 23

$t^x_k(t) = G_k s^x_k(t) - G_k$ Integral $s^x_k(t)\,dt$ Eq. 16.2.2-7 23

$t^x_L(t) = G_k s^x_L(t) - G_k$Integral $s^x_L(t)\,dt$ Eq. 16.2.2-8 23

$[E(l)S(l)](t) = \{\,[E(l)S^x(l)] - [E(l)S^{x-1}(l)]\}/\,Dx/V.$ Eq. 16.2.2-9 23

$t^x_k(t) = G_k s^x_k(t) + G_k s^{D\,x}{}_k(t)(V/Dx) - G_k$ Integral $s^x_k(t)\,dt$ Eq. 16.2.2-10 24

$t^x_L(t) = G_k s^x_L(t) + G_L s^{D\,x}{}_L(t)(V/Dx) - G_k$ Integral $s^x_L(t)\,dt$ Eq. 16.2.2-11 24

$n^x_k(t) = 1 - G_k t_k(t)$ Eq. 16.2.2-12 24

$n^x_L(t) = 1 - G_L t_L(t)$ Eq. 16.2.2-13 24

$w^x_k(t) = C_k \text{ Ln } n^x_k(t) = 0 - \text{Ln}[G_k t^x_k(t)]$ Eq. 16.2.2-14 24

$w^x_L(t) = C_L \text{Ln } n^x_L(t) = 0 - \text{Ln}[G_L t^x_L(t)]$ Eq. 16.2.2-15 24

Matrix form of The Vision Equations for a tetrachromat Eq. 16.2.4-1 27

$R(t) = m_{RU}(w_U) + m_{RS}(w_S) + m_{RU}(w_M) + m_{RU}(w_L)$ Eq. 16.3.2-1 31

$R(t) = m_{RU}(C_U \text{ Ln } n_U\,) + m_{RS}(C_S \text{ Ln } n_S\,) + m_{RM}(C_M \text{ Ln } n_M\,) + m_{RL}(C_L \text{ Ln } n_L\,)$ Eq. 16.3.2-2 31

$R(t) = \S\, m_{Rk}(C_k \text{ Ln } n_k\,) = \S\, m_{Rk}(C_k \text{ Ln } G_k\, t_k)$ Eq. 16.3.2-3 31

$R(t) = \S\, m_{Rk}(C_k \text{ Ln } G_k$ Integral $E(l)\,S(l)O(l)\,R_k(l)\,dl) +$
$m_{RL}(C_L \text{ Ln } G_L\,[\text{Integral } E(l)\,S(l)O(l)\,R_L(l)\,d\,l)]^2)$ Eq. 16.3.2-4 . 31

$R(t) = \S\, m_{Rk}\,C_k \text{ Ln } G_k \text{ modulation}(t) + m_{RL}(2C_L \text{ Ln } G_L \text{ modulation}(t)$ Eq. 16.3.2-5 32

$O(t) = \pm(m_{OU}w_U - m_{OS}w_S)$ Eq. 16.3.4-1 40

$P(t) = \pm(m_{OS}w_S - m_{OM}w_M)$ Eq. 16.3.4-2 40

$Q(t) = \pm(m_{OM}w_{M_} m_{OL}w_{L)}$ Eq. 16.3.4-3 40

$O(t) = m_{OU}C_U \text{ Ln } G_U\, s_U(t) - m_{OS}\,C_S \text{ Ln } G_S\, s_S(t)$ Eq. 16.3.4-4 41

$P(t) = m_{PS}C_S \text{ Ln } G_S\, s_S(t) - m_{PM}\,C_M \text{ Ln } G_M\, s_M(t)$ Eq. 16.3.4-5 41

$Q(t) = m_{QL}C_L \text{ Ln } G_L\, s_L(t) - m_{QM}\,C_M \text{ Ln } G_M\, s_M(t)$ Eq. 16.3.4-6 41

$O(t) = \text{Ln}[G_U\, s_U(t)/\,G_S\, s_S(t)]$ Eq. 16.3.4-7 41

$P(t) = \text{Ln}[G_S\, s_S(t)/\,G_M\, s_M(t)]$ Eq. 16.3.4-8 41

$Q(t) = \text{Ln}[G_L\, s_L(t)/\,G_M\, s_M(t)]$ Eq. 16.3.4-9 41

$O(t) = \text{Ln}[G_U\, s_U \text{ mod}_U(t)/\,G_S\, s_S \text{ mod}_S(t)] = \text{Ln}[\text{mod}_U(t)/\text{mod}_S(t)]$ Eq. 16.3.4-10 41

$P(t) = \text{Ln}[\text{mod}_S(t)/\text{mod}_M(t)]$ Eq. 16.3.4-11 41

$Q(t) = \text{Ln}[\text{mod}_L(t)/\text{mod}_M(t)]$ Eq. 16.3.4-12 41

$P(t) = m_{PS}C_S \text{ Ln } G_S$ Integral $E(l\,)\,S^x(l\,)O(\,l)\,R_S(\,l)\,dl$
$- m_{PM}\,C_M \text{ Ln } G_M$ Integral $E(l\,)\,S^x(l\,)O(\,l)\,R_M(\,l)\,dl$ Eq. 16.3.4-13 42

$Q(t) = m_{QL}C_L \text{ Ln } G_L$ Integral $E(l\,)\,S^x(l\,)O(l\,)\,R_L(\,l)\,dl$
$- m_{QM}\,C_M \text{ Ln } G_M$ Integral $\int E(l\,)\,S^x(l\,)O(\,l)\,R_M(\,l)\,dl$ Eq. 16.3.4-14 42

$P(t) = m_{PS}C_S \text{ Ln } Sp(l)R_S(l) - m_{PM}\,C_M \text{ Ln } Sp(l)R_M(l)$ Eq. 16.3.4-15 42

$Q(t) = m_{QL}C_L \text{ Ln } Sp(l)R_L(l) - m_{QM}\,C_M \text{ Ln } Sp(l)R_M(l)$ Eq. 16.3.4-16 42

$P(t) = m_{PS}C_S \text{ Ln } Sp(l)R_S(l)\ \&\ Q(t) = 0.00$ Eq. 16.3.4-17 42

$P(t) = 0.00\ \&\ Q(t) = m_{QL}C_L \text{ Ln } Sp(l)R_L(l)$ Eq. 16.3.4-18 42

$P(t) = Q(t) = -\,m_{PM}\,C_M \text{ Ln } Sp(l)R_M(l)$ Eq. 16.3.4-19 42

$P(t) = H_{PQ} + m_{PS}C_S \text{ Ln } [H_{SQ} + G_S\, s_S(t)] - m_{PM}\,C_M \text{ Ln}[H_{MQ} + G_M\, s_M(t)]$ Eq. 16.3.4-20 43

$Q(t) = H_{QQ} + m_{QL}C_L \text{ Ln}[H_{LQ} + G_L\, s_L(t)] - m_{QM}\,C_M \text{ Ln}[H_{MQ} + G_M\, s_M(t)]$ Eq. 16.3.4-21 43

$t \bullet (dq/dt) + q = Q_2 \bullet q_f$ Eq. 16.4.1-1 49

Complete Photodetection/De-excitation equation Eq. 16.4.1-2 51

The delay term of the Photodetection/De-excitation equation Eq. 16.4.1-3 52

The amplitude function of the complete P/D Equation Eq. 16.4.1-4 55

Complete Photodetection/De-excitation equation Eq. 16.4.1-5 58
First Derivative of the P/D Equation Eq. 16.4.1.6 59
Eq. 16.4.2-2 68
Eq. 16.4.2-4 68
$K(E_v,T, l,Dl)$ = Eq: 16.6.1-1 95

LIST OF FIGURES

Figure 16.1.2-1 Formation of candidate conformal color spaces 9
Figure 16.2.1-1 The luminance, chrominance and appearance channels of the eye of tetrachromats 13
Figure 16.2.1-2 Important signaling channels, sources of non-linearity and asymmetry in the visual process of animals 16
Figure 16.2.1-3 The operational amplifier architecture of the photoreceptor cell 20
Figure 16.3.2-1 The luminous efficiency function of the aphakic human eye 32
Figure 16.3.3-1 Calculation of the absorption of the physiological optics 34
Figure 16.3.3-2 The losses associated with the physiological optics of the eye 36
Figure 16.3.3-3 The complete luminous efficiency function for the human eye 39
Figure 16.3.4-1 Response of the chrominance channels versus an equal integrated flux monochromatic stimulus 44
Figure 16.3.4-2 Color discrimination versus wavelength for two tetartanopes 45
Figure 16.4.1-1 The fundamental mechanisms of photoexcitation/de-excitation 51
Figure 16.4.1-2 The delay as a function of absorbed irradiance in the P/D Equation 52
Figure 16.4.1-3 Signal delay in different species 53
Figure 16.4.1-4 A family of parametric curves for the P/D Equation 57
Figure 16.4.1-5 A comparison of the complete P/D Equation versus the data of Baylor, et. al 60
Figure 16.4.1-6 A comparison of the complete P/D Equation versus the ERG data of Smith & Lamb . . . 61
Figure 16.4.2-1 Overall adaptation block diagram 64
Figure 16.4.2-2 Expanded PC-RPE interface showing electrical analog 66
Figure 16.4.2-3 Adaptation amplifier block diagram. 67
Figure 16.4.2-4 Idealized circuit schematic representing the vascular, electrostenolytic and electronic elements 67
Figure 16.4.2-5 Predicted collector voltage of adaptation amplifier 69
Figure 16.4.2-6 The theoretical dark adaptation worksheet 71
Figure 16.4.2-7 Dark adaptation curve of a human observer 73
Figure 16.4.2-8 Dark adaptation at different distances from center 74
Figure 16.4.2-9 REWORK Theoretical light adaptation characteristic. 77
Figure 16.4.5-1 Fundamental circuit of the signal projection 78
Figure 16.4.5-2 CR Topology of the brain's input and transfer circuitry. EXTEND CAPTION 81
Figure 16.4.5-3 Asymmetrical operation of the signal projection codec 82
Figure 16.4.5-4 Waveforms associated with the signal projection stage. 83
Figure 16.4.5-5 Selected waveforms of the signal projection stage. 85
Figure 16.5.1-1 Top level schematic of the visual system of Chordata and the waveform Class 87
Figure 16.5.1-2 Collected waveform Classes for large signals 89
Figure 16.5.1-3 Fundamental waveforms and delay times 92
Figure 16.5.1-4 Asymmetrical operation of the signal projection codec. 94
Figure 16.6.1-1 Obtaining a Fourier Transform by optical means 99
Figure 16.6.1-2 Relationship between Nyquist frequency and Nyquist limit 102
Figure 16.6.1-3 Fourier planes of two interference patterns of the same spatial frequency 104
Figure 16.6.1-4 Fourier planes of the sum of two interfering fringe patterns after passing through a nonlinear mechanism 106
Figure 16.6.1-5 Comparison of the data of Polyak and the analyses of Yellott and Snyder & Miller . . . 107
Figure 16.6.1-6 Obtaining a Fourier Transform by optical means 108
Figure 16.7.1-1 Empirical data supporting the two-chrominance channel concept of this work 111
Figure 16.7.1-2 Action potential parameters of cat in optic tract and LGN 112
Figure 16.7.1-3 Action potential parameters for cat in the presence of a uniform background field 113
Figure 16.7.2-1 The signal and grounding plan of the human retina. 116
Figure 16.7.2-2 Signal flow within the fundamental signal path of the retina 118
Figure 16.7.2-3 {SCALE} The synthesis of an ERG based on predicted waveforms 127

(Active) SUBJECT INDEX (using advanced indexing option)

action potential 45, 79-82, 84, 88, 91-93, 110, 112, 113, 119, 123, 125, 128, 129, 131, 134
Activa 15, 18, 20, 25, 26, 43, 51, 52, 58, 60, 65-69, 75, 76, 79, 82-84, 86-88, 92-94, 117-119, 129, 131
adaptation . . . 4, 6, 10-12, 15-21, 23-25, 29, 31, 32, 39, 41-44, 48, 49, 52, 53, 59, 60, 62-77, 86, 87, 90, 91, 93-96, 101, 110, 111, 113, 114, 119, 122-127, 129, 130, 132, 133
adaptation amplifier . 11, 15-17, 19, 23, 24, 29, 39, 43, 49, 52, 53, 59, 60, 62, 64, 66-69, 75, 76, 87, 90, 91, 93, 94, 101, 113, 119, 122, 125, 126, 130, 133
amplification 2, 12, 19, 20, 50, 66, 79, 110, 125
arborization 31
area 7 48
attention 6, 8, 28, 86, 121, 124
average velocity 53
axon segment 20, 120
axoplasm 62, 79, 87, 88
a-wave 50, 51, 58, 61, 117, 120, 121, 126, 127, 129-132, 134-136
band gap 111
bifurcation 79
Black Body 30
bleaching 52, 72, 73
broadband 95, 96, 110
Brownian motion 63
b-wave 120-125, 130, 132, 134-136
C.I.E. 5-7, 30, 31, 33, 39, 95
calibration 5
cerebellum 47, 48
cGMP 129
CIE 6, 7, 40
CIE 1960 . . . . . 6, 7
CIE 1976 . . . . . 6, 7
class A 50, 87, 91, 92, 119, 121, 127, 128, 130, 132
class B 59, 87, 93, 125
colliculus 40
commissure 90
computation 70, 98
computational 120, 134
confirmation 2, 86
consonance 64
continuum 45, 123
cross section 52, 56, 61
cross-section 4, 5, 22, 61, 131, 133
cut-in 130
dark adaptation 20, 64, 68-75, 96, 123-125, 129, 132, 133
data base 72, 132
database 10, 45, 106
decoder 45
dendrolemma 65, 66
determinants 75
diode 16, 20, 24, 25, 43, 62, 65, 66, 76, 112, 121
dynamic range 4, 17, 20, 39, 44, 60, 91, 95
eccentric cell 53, 134
electrostenolytic process 65, 66, 69, 75, 94, 119, 122, 130
ERG 3, 50, 52, 55, 59, 61, 66, 88, 109, 114-117, 119-136
evolution 33
expanded 17, 22, 27, 31, 40, 41, 43, 57, 66, 72, 76, 133
exposine 67, 68, 71-73
external feedback 11, 20, 67
feedback 4, 10, 11, 16, 19, 20, 23, 24, 66, 67, 112
Fourier transform 12, 97-100, 107, 108
free-running 114
Gaussian 46, 50, 61, 129
glutamate 75

group velocity 85, 129
g-protein 129
Hodgkin solution 57, 59
hole 87, 126
homogeneous 38
hydronium 54, 65, 75
hydronium liquid crystal 75
internal feedback 4, 11, 19, 20, 66
interneuron 79, 83, 84, 88
inverting 20, 25, 26, 84, 94, 118
Kirchoff's Laws 12
latency 50, 61, 129, 130, 134, 135
lateral geniculate 46, 111
light adaptation 10, 21, 64, 68, 69, 72, 76, 77, 96
Limulus 49, 53, 55, 62, 63, 91, 134
logistic function 121, 130
mesotopic 43, 44, 113
metabolites 75, 115
microtubule 20
midbrain 25, 47, 89, 90, 101
modulation 17, 32, 41, 64, 85, 99, 109, 113, 114
monopulse 113
monopulse oscillator 113
multi-dimensional 14, 27, 101
narrow band 5, 29, 42
neural coding 13
neurite 88
neurites 43
Node of Ranvier 79, 82, 83, 88, 92, 119
noise 5, 47, 57, 59, 60, 75, 81, 82, 85, 108, 111, 114, 120, 123, 129
non-inverting 26, 94, 118
Nyquist 100-102, 104, 106-108
Ohm's Law 12
p/2 70
P/D equation 11, 19, 21, 49, 51-63, 68, 73, 75, 76, 87, 88, 90, 91, 96, 119, 128, 129, 135, 136
parametric 10, 57
pedestal 43, 62
perceptual space 2
pgn/pulvinar 14
phase velocity 83
poditic 118, 119
POS 14, 47, 48
Pretectum 13, 14, 25, 47, 48, 64, 101, 130
protocol 109, 130
pulvinar 14, 48
Pulvinar pathway 48
quantum-mechanical 11
reading 48, 133
resonance 38, 39
retinitis pigmentosa 130, 135
ringing 124
saliency map 14, 21, 45, 48
serape 126
servo loop 64
signal-to-noise 108
signal-to-noise ratio 108
spectral colors 96
square law 16, 43
square-law 86
stage 0 33

stage 1 . . . . . 12, 14, 18-20, 22, 49, 63, 115
stage 2 . . . . . 3, 12, 19, 20, 23-25, 29, 77, 101, 115
stage 3 . . . . . 3, 11, 14, 15, 28, 29, 40, 43-45, 77, 101, 109
stage 4 . . . . . 3, 14, 114
stage B 12
Standard Observer 7, 33
stellate 15, 28, 44, 45, 78-80, 84, 88-90, 92, 111, 123
stellate cell 80, 84, 90, 92
Stiles-Crawford 5, 64
stress 33, 70, 88, 106, 111
superior colliculus 40
synapse 25, 78, 83, 84, 88, 118, 119
tetrahedron 7
threshold 4-6, 17, 47, 48, 64, 73, 75, 76, 81, 84, 90-95, 104, 114, 135
topography 115, 116
topology 81, 115, 123, 132
transduction 16, 19-22, 48, 61, 72, 76
translation 16, 19, 20, 22, 43
tremor 23, 24, 29, 48, 101
type 2 88
type B 51
V2 75
visual cortex 84, 92
xxx 5, 15, 23, 24, 29, 43, 47, 51, 55, 61, 62, 65, 67, 78, 85, 94-96, 111, 114, 120, 123, 129, 135
[xxx 8, 24, 33, 40, 62, 64, 67, 77, 109, 111, 128, 142

(Inactive) DEFINITIONS INDEX (Use individual marks)

Principle of Univariance
Retinal illuminance
Transport delay
net photoreceptor