Visual latency and brightness:  
An interpretation based on the responses of rods and ganglion cells in the frog retina

KRISTIAN DONNER
Department of Zoology, University of Helsinki, SF-00100 Helsinki, Finland

(RECEIVED August 24, 1988; ACCEPTED March 8, 1989)

Abstract
Rod and cone photoresponses in a variety of species have been accurately described with linear multistage filter models. In this study, the response latency and initial coding of intensity at two higher levels of visual processing are related to such photoreceptor responses. One level is the retinal output (spiking discharges from frog ganglion cells, based on experimental data reported here), the other is the perceptual level in humans (psychophysical latency and brightness functions, based on data from the literature). Photoreceptor responses are described with the “independent activation” model of Baylor et al. (1974). The intensity dependence of the early ganglion cell discharge, its latency and initial impulse frequency, is shown to follow from such a waveform, assuming that 1) latency $L = t + D$, where $t$ is the time it takes for the rod response linearly summed over the ganglion cell’s receptive field to reach a criterion amplitude, and $D$ is a constant delay; and 2) the initial frequency (below saturation) is proportional to the steepness of rise of the summed rod response at time $t$. It is shown that the intensity dependences of 1) human visual latency and 2) brightness sensation, including effects of stimulus area and duration, are accounted for by the same model. The predicted functions are not power functions of intensity, but approximate such over wide ranges. Thus, a large body of psychophysical data is explained simply by the waveform of photoreceptor responses.

Keywords: Photoreceptors, Impulse frequency, Neural intensity coding, Psychophysical power law

Introduction
All visual signals are initially filtered through the photoreceptors. The object of the present study is to relate psychophysical data on visual latency and brightness, as well as corresponding data from frog ganglion cells, not to stimulus intensities as such, but to the photoreceptor responses elicited by these stimulus intensities. It is shown that a wealth of psychophysical data can be explained by assuming a very simple linear dependence on photoreceptor responses.

The responses of vertebrate rods and cones to flashes and steps of light have been thoroughly investigated (e.g. Baylor & Fuortes, 1970; Penn & Hagnis, 1972; Baylor & Hodgkin, 1973; Baylor et al., 1974; Fain, 1976; Schwartz, 1976; Cervetto et al., 1977; Baylor et al., 1979; Baylor et al., 1984), and it seems likely that phototransduction is similar in different species. At every level in the visual system, the earliest response to the onset of light could be fairly directly determined by phototransduction, because any feedback will be delayed in relation to the primary signal. It should then be possible to formulate a simple model relating the rising receptor response to the early response at higher levels. Moreover, basically the same relations should hold at different levels and for different species. Here, such a unified model is developed and applied to the discharges of frog ganglion cells and to human psychophysics in the dark-adapted state.

First, the impulse discharges of ganglion cells in the dark-adapted frog retina are considered. It is shown that the early response at the retinal output—its latency and initial impulse frequency—bears a straightforward linear relation to the rising rod response. Then, observing that frog ganglion cells closely replicate power functions found to describe, in human vision, the intensity dependence of latency (Vaughan & Hull, 1965; Vaughan et al., 1966; Alpern, 1968; Mansfield, 1973; Burkhardt et al., 1987) and brightness (e.g. Hopkinson, 1956; Stevens, 1961; Aiba & Stevens, 1964; Marks & Stevens, 1966; Barlow & Verriolo, 1976), it is demonstrated that these psychophysical functions can be similarly generated from receptor responses. Although the resulting functions are not power functions of stimulus intensity, they can be approximated with such, but with different exponents over different ranges of stimulus flux.
Methods

Materials and recording

Materials

Frogs (Rana temporaria) were caught in the autumn in southern Finland and stored in basins at 3-4°C without feeding. Before dissection, the frog was allowed to warm gradually over some 18 h to about 15°C in a dark chamber. Dissection was carried out under dim red light and the preparation was then left to dark adapt for two more hours in the specimen chamber at 10-12°C, which was the temperature of all the experiments. (It is to be noted that 3-15°C is approximately the temperature range naturally encountered during 24 h by a Finnish frog during much of its active season, in spring and autumn (cf. also Donner et al., 1988). Experiments done in late spring and early summer on frogs caught breeding in May and then maintained at constant temperature (10-12°C) indicated no differences relevant to the present work.)

Extracellular ganglion cell recordings

Spike responses from single ganglion cells were recorded with extracellular microelectrodes in the eyecup preparation. Dissection, recording, stimulation, and light calibration techniques have all been described elsewhere (Donner & Reuter, 1968; Bäckström et al., 1978; Copenhagen et al., 1987). Stimulus intensities (I) are here expressed in units of photons emitted per rod per second, denoted Rh s⁻¹. This can be converted into stimulus flux (Φ) (isomerizations/s) within a given area by observing that there are, on average, 15,700 rods/mm² in the retina in Rana temporaria (Hemmälä & Reuter, 1981).

The isolation of units in the experiments was always such that one could be sure that all spikes came from one single cell. The preparation and recording were usually very stable: latency and frequency functions could remain entirely reproducible over as much as 10 h of recording from the same cell.

Response measures

The ganglion cells of classes 1-3 are practically silent in darkness: their maintained discharge is usually less than one spike per minute (Aho et al., 1987). When stimulated with flashes or steps of light, they respond with distinct spiking discharges, which are cut off more or less abruptly depending on the strength of the lateral inhibition activated. Yet the early part of the response to not-too-large test spots in class 3 cells is independent of lateral inhibition, consisting of a burst of spikes whose latency decreases and (presaturation) frequency increases monotonically with stimulus intensity (Donner, 1981). In class 1-2 cells, even the initial firing rate may be influenced by some “local” type of inhibition. The present work is concerned only with responses undisturbed by inhibition, and the treatment of initial impulse frequency will be wholly based on class 3 cells. The class 3 ganglion cells of the frog retina seem to form a complete subsystem with full coverage of the visual field (Mecke et al., 1989), well suited, e.g. for conveying the brightness information to the brain. The latency considerations apply to class 1-2 cells as well. For cell classification, see Bäckström and Reuter (1975) and Donner and Grönholt (1984).

Impulse frequency was taken as the reciprocal of the time interval between the first and the fourth spike, multiplied by 3.

Using the first four spikes gives the readings reasonable statistical stability while still providing a metric close to an “instantaneous” frequency (cf. also Donner, 1987a). Latency was taken as the time from stimulus onset to the middle of the first spike. Usually, the values given are means of three trials. The determination of 50% response thresholds and of the summation area of the receptive-field center (top-hat approximation) was done as described by Donner (1987b). Test spots much smaller than the summation area are referred to as “small” (diameters 50 or 110 μm), whereas “large” spots (diameters usually 530 or 800 μm) fully cover the summation area. All results refer to On-type responses (i.e. responses to On steps or flashes of light). “Flashes” had 17-ms duration, and “On steps” were, in fact, also square-wave pulses, never longer than 8 s.

Recording of rod responses in the aspartate-treated retina

To illustrate the relation of ganglion cell responses to rod responses in the same species at the same temperature, rod responses were recorded as transretinal mass potentials across the isolated, aspartate-treated frog retina (see Sillman et al., 1969). In this type of recording, each response is the ensemble average of responses from thousands of rods. Thus, it is possible to obtain responses with high signal/noise ratios over wide intensity ranges within a short time, since it is usually not necessary to average several responses. (It will be noted from Fig. 1 that a high signal/noise ratio is required, as the response amplitudes which concern us here are only about 1% of the maximum response.) The early rise of the mass response is essentially undistorted by glial currents (Donner & Hemmilä, 1985, Donner et al., 1988). Only such preparations were used where the dimflash response had a time-to-peak less than 4 s, i.e. within the range of times-to-peak recorded intracellularly from rods in the same species at 10-12°C (see Discussion).

Dissection, recording, stimulation, and light calibration were as described by Donner and Hemmilä (1985) and Donner et al. (1988). The high-frequency roll-off of the d.c. amplifier was from 130 Hz, high enough to ensure good reproduction of the responses considered here (see Fig. 1). The rising phases of photoresponses were recorded on fast time scales chosen to give sufficient resolution to each response on a Tektronix 5103N storage oscilloscope, photographed and manually digitized from the pictures.

Analysis

The “independent activation” model for rod responses

The analysis of early response parameters in this work is based on the modeling of rod responses with the independent activation model introduced by Baylor et al. (1974), one variant of models based on a chain of linear low-pass filters as first used by Fuortes and Hodgkin (1964) to describe responses in Limulus ommatidia. Such models have been successfully used to fit intracellularly recorded voltage responses from cones or rods (Baylor et al., 1974; Schwartz, 1976; Cervetto et al., 1977) and current responses from single rods (Baylor et al., 1979, 1983, 1984; Lamb, 1984; Schnapf, 1983) in various species, as well as the rising phase of mass rod responses from Rana temporaria (Donner et al., 1988). Considering (as here) the early rise of responses, it matters little what exact formulation is
used. In the “independent activation” variant, the response to a step of dim light is given by

\[ R_s(I, t) = aI(1 - e^{-t/\tau})^n, \]  

and the response to an “infinitely” brief dim flash by

\[ R_f(I, t) = aIt\tau^{-1}e^{-t/\tau}(1 - e^{-t/\tau})^{n-1}, \]

where \( a \) is sensitivity, \( t \) is time from stimulus onset, \( \tau \) is the largest (in the arithmetic sequence) of the time constants, \( t_f \) is flash duration, and \( n \) is the number of stages in the activation chain (Baylor et al., 1974, eqns. 40 and 41). In the present work, the term step of light will be used for any light pulse which is so long that increasing its duration no longer affects the response components under study.

Equations (1) and (2) hold only for responses to low-intensity stimuli (“linear range responses”). To describe full responses to higher stimulus intensities, other effects such as saturating nonlinearities would have to be included. These are, however, unimportant for the early rise of the responses even to high intensities (cf. Baylor et al., 1974). From eqns. (1) and (2), it further follows that the very earliest rise is a power function of time:

\[ R_s(I, t) = aIt/\tau^n, \]
\[ R_f(I, t) = aIt\tau^{-1}n(t/\tau)^{n-1}, \]

as \( t \) approaches zero.

The time constant \( \tau \) bears simple relations to two other important kinetic parameters (see Donner (1985), Appendix to part IV):

1. The integration (or summation) time \( t_i \), defined \( t_i = I_f / I_s \).
   For a rod, \( I_f \) and \( I_s \) are the stimulus intensities of a flash and a step stimulus, respectively, eliciting responses of a given small criterion amplitude (cf. Baylor & Hodgkin, 1973); for a ganglion cell, they are the respective threshold intensities (cf. Donner, 1987b). Then
   \[ t_i = \tau(n/(n-1))^{n-1}. \]

Equation (5), together with the definition of \( t_i \), provides a means of calculating the threshold difference (log \( I_f / I_s \)) of a flash and a step stimulus from the parameters \( n \) and \( \tau \).

2. The time-to-peak \( t_p \) of the receptor response to a flash is related to \( \tau \) as
   \[ t_p = \tau(\ln n). \]

Power-function fits to latencies

From the mid-1950’s to the early 1970’s, psychophysical data were predominantly fitted with power functions (see e.g. Stevens, 1961). As the relation of that work to the model proposed here must be clarified, latencies were analyzed also as power functions of intensity. Ganglion cell latency \( L \) is construed as the sum of an asymptotical delay \( L_\infty \) and an intensity-dependent portion \( L - L_\infty \). The intensity-dependent portion is then fitted with a negative power function of stimulus intensity \( I \) (see Vaughan et al., 1966; Mansfield, 1973a):

\[ L - L_\infty = kI^{-\beta} \]

where \( \beta \) is the exponent and \( k \) is a proportionality constant. This phenomenological description can also be used as a tool for assessing the similarity of rod, ganglion cell, and psychophysical latency functions. On log-log coordinates, the power function is a straight line:

\[ \log(L - L_\infty) = -\beta \log I + \log k \]

The \( \beta \) and \( L_\infty \) values giving an optimal fit of eqn. (8) to a set of latencies \( L \) was found by an iterative computer program maximizing the Pearson product-moment correlation coefficient \( r^2 \) (leaving least residual variance).

The most important symbols used

\( L \) = the latency of the ganglion cell’s spiking discharge
\( L_\infty \) = the asymptotic latency obtained when power functions are fitted to ganglion cell latencies (eqn. (8))
\( I \) = receptor “latency,” i.e. the time to a criterion amplitude of receptor responses (recorded or calculated according to eqns. (1) or (2))
\( L_\infty \) = the asymptotic latency obtained when power functions are fitted to the times-to-criterion of recorded or theoretical receptor responses
\( D \) = hypothetical transmission delay
\( d \) = \( L_\infty - L_\infty \), an estimate of \( D \)
\( \beta \) = the exponent of a power-function fit according to eqn. (8)
\( \tau \) = the largest of the time constants in the independent activation model
\( n \) = the apparent number of stages in the filter chain (eqns. (1–6))
\( \phi_T \) = the light flux needed for a threshold response of a ganglion cell, i.e. (spot area) times (threshold intensity) for a small test spot, or (ganglion cell summation area) times (threshold intensity) for a large test spot.

Results

Requirements for a model relating the early response of rods and ganglion cells

The general relation between rod responses and the early ganglion cell discharge

The quantitative analysis of ganglion cell discharges (as well as human psychophysics) in this study is based on modeled photoreceptor waveforms such as have been found to fit both rod and cone responses in a wide variety of vertebrate species (see Methods). To begin with, however, it is useful to inspect rod and ganglion cell responses recorded from one and the same species at one and the same temperature. This will give a qualitative idea of their relation. Such comparable response families from rods and a ganglion cell in the frog retina are shown in Fig. 1. On one hand, it illustrates which aspects of the rod response are relevant for the formation of the early discharge of the ganglion cell. On the other hand, it illustrates the appli-
cation of a linear multistage filter model to the rising phase of rod responses in *Rana temporaria* at 11°C.

The dotted curves in Fig. 1A show the rising phases of rod responses to step stimulation, recorded over a 5-log-unit range of stimulus intensities at 0.5-log-unit intervals. The abscissa gives time from stimulus onset, the ordinate relative response amplitudes (as fractions of the saturating response amplitude $U_{\text{max}}$). The continuous curves trace eqn. (1) with $n = 4$, which is the number of stages generally found appropriate for current responses from perfused amphibian rods (Baylor et al., 1979, 1983; Lamb, 1984; Schnapf, 1983; Donner et al., 1988). The model is seen to provide an acceptable fit. (Note that the curves have not been individually fitted, but the whole family is based on a single set of parameters, see figure legend.) One particular implication (cf. eqn. (1)) is that the segments of rod responses in the figure are linear in intensity over the whole range shown.

The lower panel (Fig. 1B) shows the spike responses of a class 3 ganglion cell to step stimulation with the same light intensities. Here, each dot represents one nerve impulse and each row of dots one discharge in response to the stimulus intensity indicated on the ordinate. (Three discharges are shown at each intensity.) The abscissa again gives time from stimulus onset, but in this case stimulus onset (the zero of the time axis) has been displaced to lie 0.2 s earlier than in panel A. The vertical lines are drawn to indicate the mean time of response onset at each intensity.

When relating Fig. 1A to 1B in detail, it must be borne in mind that strong quantitative inferences are impermissible. There can be no absolute correspondence between amplitudes...
of mass receptor responses (which depend on the recording configuration) and ganglion cell responses. Moreover, ganglion cell responses depend on the spatial extent of the stimulus (see below); thus the precise intensities are relevant for the illuminated rods, but not directly for the ganglion cell.

These reservations cannot obscure two important conclusions of a qualitative nature. Firstly, all of the ganglion cell discharges commence at roughly the same (very small) amplitude of the corresponding receptor response, regardless of stimulus intensity. Put somewhat differently, the time to a small criterion amplitude of the receptor response and the latency of the ganglion cell discharge shorten with intensity in a similar manner. (Remember, though, that the relative displacement of the rod and ganglion cell time axes in the figure amounts to having subtracted a constant "transmission delay" of 0.2 s from all of the ganglion cell latencies: see below). Secondly, it is obvious that the ganglion cell discharge is associated with the early rising phase of the receptor response. The discharge is generally over and done with before the rod response has reached more than a fraction of its final amplitude.

**Flux-dependence of the early ganglion cell response**

Frog ganglion cells sum rod signals over receptive fields encompassing large numbers of rods, varying from about one hundred up to several thousands in different cells (Bäckström & Reuter, 1975; Donner, 1981; Donner & Grönholm, 1984; Aho et al., 1987). Thus, a model for the early discharge must incorporate a spatial summation factor. At threshold, this summation is known to be linear, so that response thresholds are determined by the light flux absorbed within the receptive-field center (Hartline, 1940). Figure 2 indicates that this is true of latency and initial impulse frequency even at fairly high stimulus intensities (cf. also Cleland & Enroth–Cugell, 1968; Donner & Grönholm, 1984).

Pairs of latency and impulse-frequency functions have here been extracted from the On responses to a small (triangles) and a large (circles) test spot centered on the receptive field. In Fig. 2A, latencies are plotted against log intensity in the form of reciprocal latencies ("response speed"). This presentation has the convenient property that nearly straight lines result over a considerable range of intensities. It is then easily seen that over long stretches, the small-spot function replicates the large-spot function, only it is relatively displaced to higher log intensities. The distance between the functions on the log I axis turns out to be about the same as the difference in log response thresholds of the two spots (0.8 log units in Fig. 2). As a consequence, the latency functions will roughly coincide if plotted against the photon flux absorbed by the cell's receptive field (i.e. relatively displaced by the log threshold difference). This has been done in Fig. 2B (note that the latency ordinate here is linear, not reciprocal).

Figure 2C and D show the same to hold (with some reservations) for the initial impulse frequency. Panel C displays log frequencies for the two spots as functions of log intensity. When plotted instead against log flux as in panel D, the frequencies fall on a single curve, before saturating at different levels. In all

![Fig. 2. The early ganglion cell response as a function of photon flux linearly summed over the receptive field. Circles: data from responses to a large spot (0.8-mm diam); triangles: data from responses to a small spot (0.1-mm diam), both centered on the receptive field (about 0.28-mm diam, encompassing about 1000 rods). 513-nm step stimulation, class 3 cell. A, reciprocal response latencies shown as functions of log stimulus intensity. Log threshold was 2.07 (Rh s⁻¹) for the large spot and 1.27 (Rh s⁻¹) for the small spot, giving the difference 0.8 log units. This is very nearly the mean distance between the two reciprocal latency lines. B, the same data as in A, but displayed as linear latencies against log stimulus flux. Log flux is given as log φ/φᵣ, where φᵣ is the threshold flux, 8 isomerizations/s within the receptive-field center. C, log initial impulse frequency (impulses/s), as function of log stimulus intensity. D, the same data as in C as function of log stimulus flux. The continuous curves in the plots on flux abscissa (B and D) are model functions described later in the text.]
10 cells where frequency functions were determined with both a large and a small spot, this was found to hold as a good first-order approximation.

Thus, the early discharge is essentially flux determined (below saturation). This interchangeability of area and intensity seems to imply that, up to the stage of final spatial summation, the signal underlying the early response is linear in intensity.

The model

The main observations from the preceding are 1) the early rod response can be described with eqn. (1) or (2) and is linear in intensity, 2) the ganglion cell response is determined by the rising phase of the rod response as indicated by Fig. 1; and 3) the signal underlying the early ganglion cell discharge is linear in intensity and sums linearly in space. In the following, these requirements are worked into a coherent model.

Latency

The ganglion cell is assumed to start firing when the receptor response linearly summed over the receptive field has reached a (very small) criterion amplitude and been transmitted up to the ganglion cell. Rod responses according to eqns. (1) and (2) have no irreducible delay. Thus, ganglion cell latency \( L \) is seen as the sum of an intensity-dependent receptor time-to-criterion \( l \), which approaches zero when the light flux approaches infinity, and a constant transmission delay \( D \), \( L = l + D \). Predicted latency functions (for step stimulation), as calculated from eqn. (1) with \( n = 3, 4, 5, \) and \( 6 \), are displayed in Fig. 3A.

Initial impulse frequency

After the first spike, every additional action potential is assumed to correspond to an additional criterion rise of the receptor response. An idealized “instantaneous” impulse frequency would then be proportional to the steepness (derivative) of the receptor response at the moment \( l \) when criterion amplitude is reached. Figure 3B displays the resulting dependence of log frequency on log relative flux (log \( \phi / \phi_T \)) as calculated from eqn. (1) for step stimulation and \( n \) values of 3, 4, 5, and 6. “Flash” functions calculated from eqn. (2) are rather similar to the “step” functions shown except that the \( n \) values then have to be interpreted as values of \( n - 1 \).

It will be noted that the model not only predicts latency and frequency functions separately, but also a specific relation between them: the crucial parameters \( n \) (number of stages) and \( \phi_T \) (threshold flux) pertain equally to both. Latencies further depend on the parameters \( \tau \) and \( D \), while the frequency function contains one free parameter specifying the proportionality between slopes of receptor responses and impulse frequency (the “gain” of the transformation: in a log-log plot the vertical position of the frequency curve). Also note that the saturation of impulse frequency does not follow from the proposed underlying function.

Psychophysical latency and brightness

Psychophysical latency is assumed to be determined in the same way as ganglion cell latency, except that \( D \) will contain additional delays. Perceived brightness is assumed to be determined as the initial impulse frequency of ganglion cells, except that brightness is nonsaturating over a much wider intensity range than the impulse frequency of frog ganglion cells.

The status of power functions

Power functions have held a place of honor in psychophysical work on latency and brightness, so it is important to clarify how the model is related to such functions. In responses to very high stimulus intensities, the time to a small criterion will be very short and the approximations expressed by eqns. (3) or (4) will hold. In that case, latency will decrease as intensity \( I \) raised to the power \(-1/n\) for step stimuli and \(-1/(n-1)\) for flash stimuli. The frequency-determining signal in response to a step of light is then given by

\[
\frac{dR}{dt_{rel}} = a n^{-1} I^{(1/n)} - I^{1/n}
\]  

(9)

because \( I \propto I^{1/n} \), i.e. frequency will grow as the power \( 1/n \) of intensity. In responses to flashes of light, it will grow with intensity raised to the power \( 1/(n-1) \).

However, these are only limiting cases. At somewhat lower stimulus intensities, the times-to-criterion will in general not be extremely short (see Fig. 1), and the full expressions for \( R_1(I, t) \) and \( R_2(I, t) \) have to be used. One may still choose to describe latency and frequency (or brightness) with power functions. The problem is that the exponent \( n \) will not be constant, but will vary with the intensity range considered. This is immediately evident for the frequency functions in Fig. 3B: the fact that the functions are curved in a log-log plot implies that there is no constant exponent.

In latency functions, the situation is very similar. This can be demonstrated by fitting power functions to sets of theoretical “step” times-to-criterion calculated from eqn. (1). Figure 4 shows (for \( n \) values 4 and 6) how the optimal \( \beta \) and \( l_\infty \) values (ordinates) will strongly depend on the flux (or intensity) range used for fitting. The abscissa gives the width of that range, while the different symbols correspond to different starting points: threshold (circles), 1 log unit above threshold (squares), and 2 log units above threshold (triangles). Consider, for example, the \( \beta \) values obtained for \( n = 4 \) (panel A). It is seen that the “expected” value \( \beta = 1/n = 1/4 \) emerges only as a limiting case when the set of stimuli used for the determination is strongly dominated by high intensities. \( \beta \) then gets monotonically larger as intensities closer to threshold are given more weight. Yet \( r^2 \) is \( > 0.99 \) for all the fits, i.e. technically they are impeccable. Similarly, \( l_\infty \), which should be \( = 0 \) for the theoretical responses, will take on finite values, approaching 0 only for intensity ranges far above threshold. (Because of this, \( L_\infty \) in eqn. (7) is a bad estimate for the transmission delay \( D \), which is more accurately estimated by \( d = L_\infty - l_\infty \).)

Application to ganglion cell latency functions

In Fig. 2B, the continuous curve giving latency as a function of log flux is a model time-to-criterion function with \( n = 4, \tau = 2.1 \text{ s}, \) and \( d = 190 \text{ ms} \). The fit is obviously acceptable, but is open to the objection that the shape of theoretical latency curves depends on two parameters, \( n \) and \( \tau \), and different combinations of the two yield rather similar curves. However, as shown by Fig. 4, it is possible to deduce \( n \) from the \( \beta \) value of a power function fit, if only the flux range that has been used for fitting
is known. This offers a procedure for estimating $n$ without involving $\tau$.

Power functions were optimized to latency data obtained from 29 ganglion cells by large-spot step stimulation over a standardized flux range covering about 5 log units, starting from about 1 log unit above threshold (to avoid the steepest range dependence of $\beta$ and $L_\infty$, cf. Fig. 4). Excellent fits ($r^2 > 0.99$) could always be achieved, the mean parameters under these conditions being $\beta = 0.27 \pm 0.01$ and $L_\infty = 167 \pm 6$ ms.

In view of the flux range used, $\beta = 0.27$ indicates $n = 5$ on average. The average transmission delay as estimated by $d = L_\infty - L_\infty$ would be 142 ms ($\tau$ has then been taken as 1.2 s, see below).

Each individual $\beta$ value was also separately converted into an $n$ value. Ten of the 29 sets of latency data were in best accordance with $n = 4$, twelve with $n = 5$, and seven with $n = 6$ or 7. With the $n$ value fixed, the optimal $\tau$ value was then determined for each set of data. The interdependence of the parameters $n$ and $\tau$ is evident from the fact that mean $\tau$ was $1.7 \pm$
The outcome of power-function fits to model times-to-criterion, illustrated for $n = 4$ in A–B and $n = 6$ in C–D. It is shown how the apparent exponent $\beta$ (panels A and C) and apparent asymptotical latency $l_\infty$ (panels B and D) will vary depending on the range of stimulus flux (or intensity) used. In each of the panels, the abscissa gives the width of the flux range used, while the three curves correspond to different starting points: threshold (circles), 1 log unit above threshold (squares) and 2 log units above threshold (triangles). The fits were made to "step" times-to-criterion. Dashed lines in A and C: the value $\beta = 1/n$ (i.e. 1/4 and 1/6, respectively), which is only asymptotically approached. Likewise, in B and D, the expected value $l_\infty = 0$ appears only as a limiting value.

0.2 s for the data with $n = 4$, $1.2 \pm 0.2$ s for $n = 5$, and $0.85 \pm 0.08$ s for $n = 6$ or 7. These $\tau$ values are all consistent with rod kinetics found in single-rod recordings from the same and related species (see Discussion).

Effect of stimulus area

The flux dependence of latencies implies that stimulus area should have no "effects" other than as a determinant of the photon flux on the receptive field. It is then important to realize that a power-function analysis easily engenders a spurious area-effect. For example, if the large- and small-spot latencies from Fig. 2A–2B are fitted with power functions up to log $I = 4$, quite different exponents are obtained: $\beta$(large) = 0.30, $\beta$(small) = 0.37 ($r^2 > 0.99$). This difference is mainly explicable by the range dependence of $\beta$: when a fixed upper intensity limit is applied, the flux range used for fitting is narrower for the small than for the large spot. In fact, both $\beta$ values are best compatible with $n = 4$.

On the other hand, the asymptotic latencies for the same data remain different even after correction for the range dependence: $d$(large) = 190 and $d$(small) = 230 ms. In 10 cells, this difference averaged $25 \pm 6$ ms. This indicates a small but real deviation from the flux law. It appears to be connected with the fact that either stimulus flux or stimulus intensity can be equalized, but not both at the same time. When the ganglion cell sees equal fluxes from the two test spots, more distal stages with summation areas smaller than both spots will only see unequal intensities. Hence, saturation of such stages will affect small-spot functions at lower flux levels than large-spot functions (cf. the saturation of impulse frequencies at different levels in Fig. 2C–2D).

Effect of stimulus duration

The model predicts that if the "step" latencies of a cell fit a function with parameters $n$ and $\tau$, generated from eqn. (1), the "flash" latencies of the same cell should follow a function with the same parameters, but based on eqn. (2). This prediction was tested in 5 cells, where latency functions were recorded both with On steps and 17-ms flashes of light. Figure 5 displays the results of one such experiment in terms of reciprocal latencies (a presentation similar to Fig. 2A). The step data (circles) was best compatible with $n = 5$ (as deduced from the best-fitting power function, see above), whereby the optimal fit had $\tau = 1.4$ s and $d = 158$ ms. A "flash" curve was then computed for the same parameters, with no reference to the experimental flash data and introducing no arbitrary parameters. (The log threshold difference log$I_f/I_s$ was predicted by means of eqn. (5).) This model function is seen to provide an acceptable fit to the flash latencies (crosses) in Fig. 5.

Application to the impulse frequency of ganglion cells

If the response threshold and the parameter $n$ are known, the impulse-frequency function (below saturation) is uniquely predicted by the model, save for a gain factor (vertical position in a log–log plot). In Fig. 2D, the continuous line which describes the frequency data rather well up to saturation is the model "step" function for $n = 4$, the value obtained from the latencies in Fig. 2B. Greater resolution, as well as a test of the effect of

![Fig. 5. Effect of stimulus duration on latency: the latencies of responses to a step (circles) and a 17-ms flash of light (crosses) plotted as reciprocal latencies against log stimulus flux. The 513-nm test spot (0.3-mm diam) just covered the receptive-field center. The continuous lines give the reciprocal of model latencies calculated as described in the text. Note how the latency difference between step and flash gets smaller as intensity is raised, to finally disappear (here at $\phi \approx 2 \times 10^7$ isomerizations/s). At that point the critical duration for latency is exactly equal to the duration of the flash, 17 ms.](image-url)
stimulus duration on impulse frequencies, is provided by the experiment presented in Fig. 6. Step responses (circles) were recorded at denser-than-usual (0.2 log unit) intervals and flash responses (crosses) at the usual 0.5-log-unit intervals. Both latency and frequency functions are shown, so that their relation can be appreciated.

The step latencies were again used for estimating the parameters \( n, \tau, d, \) and \( \log L \) (yielding the step curve in Fig. 6A). The flash-latency function was computed strictly theoretically from these parameters as described earlier (flash curve in Fig. 6A). Then the appropriate model functions for impulse frequencies were plotted in Fig. 6B. Here, the vertical positioning of the pair of curves is the single arbitrary parameter. The curves clearly provide a good fit up to saturation, as was consistently found in all the 14 cells where initial impulse frequencies were thus analyzed.

**Application to psychophysical latency and brightness**

**Latency**

The latency functions of frog ganglion cells and human psychophysics are strikingly similar: in both cases, power functions can be used for relating reducible latency to intensity, and the exponent values encountered largely overlap. It is then rather a trivial observation that, given free choice of parameters, the model will be able to describe the psychophysical data reasonably well. However, a more exacting test of its explanatory power is to what extent the effects of stimulus area on human latency functions can be accounted for by a common underlying flux function. This would seem possible, since Alpern et al. (1970a,b) showed that the rising signal of both rods and cones is linear in intensity and sums linearly in space even in psychophysical experiments.

Mansfield (1973a), in a thorough study of human reaction time, reports latency exponents \( \beta = 0.49 \) for a small (0.05 deg) foveal stimulus, but \( \beta = 0.31, 0.32 \) for large (2 and 4 deg) stimuli. The parameters needed for the construction of the hypothetical common underlying function are available from Mansfield’s Fig. 3 and Table 2. The asymptotic latencies \( L_\infty \) were 219 ms (small) and 188 ms (large). The 0.05-deg stimulus had been presented over a 3.5-log-unit intensity range from threshold, and from Fig. 4 of the present work it is seen that 0.49 is close to the value predicted for \( n = 4 \) over this range (the exact prediction being \( \beta = 0.47 \)). A model curve with \( n = 4 \) fitted to the 0.05-deg data requires \( \tau = 100 \) ms. The estimated \( L_\infty \) would (over the 3.5-log-unit range in question) contain 0.1 \( r_\text{ms} \) spurious \( L_\infty \) (see Fig. 4B), leaving \( d = 219 - 10 = 209 \) ms. Furthermore, the 0.05-deg threshold was 1.9 log units higher than the lowest large-spot (2 deg) threshold, suggesting that the effective flux delivered by the 2-deg and 4-deg spots was at least 1.9 log units larger than that of the 0.05-deg spot at the same stimulus intensity.

Figure 7 is a replot of Mansfield’s Fig. 3. Panel A shows the reducible latencies \( L - L_\infty \) for foveal stimuli of 0.05 deg, 2 deg, and 4 deg on logarithmic axes quite as in the original figure. Panel B shows the full latencies \( L \) plotted against log intensity. Panel C, finally, shows the corrected reducible latencies \( L - d \) plotted on a log flux abscissa. This entails the following relative displacements: 1) flux correction: the 0.05-deg data have been shifted leftward. 1.9 log units relative to the 2-deg and 4-deg data (see above); and 2) the latencies are \( \text{per definitionem} \) relatively shifted vertically by the difference in their \( d \) values. When presented in this form, the data are seen to be consistent with the common underlying function proposed (continuous line, model with \( n = 4 \) and \( \tau = 100 \) ms).

Thus, the model can, in the main, account for the reported dependence of power-function slopes on stimulus area. Small-spot data simply tend to be collected over lower and narrower ranges of stimulus flux than large-spot data. In addition, however, there is a small dependence of the corrected asymptotic latency \( d \) on stimulus area, which is not predicted by the model.

**Brightness**

Marks and Stevens (1966) published 24 individual brightness functions, which they had fitted with power functions of intensity according to the equation:

\[
\log \psi = \beta \log(l - I_T) + \log k, \tag{10}
\]

where \( \psi \) is brightness, \( k \) is a proportionality constant, \( \beta \) is the power, and \( I_T \) is threshold intensity. In the present work, each of the 24 sets of data was reconsidered for an alternative fit with template curves from Fig. 3B. These templates provided fits which were at least as good as those of eqn. (10). The optimal \( n \) values were: one function with \( n = 2 \), eight with \( n = 3 \), nine with \( n = 4 \), and six with \( n = 5 \) or 6. Thus, the dispersion is similar to that found among frog ganglion cells, but the mean is displaced to somewhat lower \( n \) values.

The steepness of brightness functions is known to vary with both the area and the duration of the stimulus. Again, it is a
Fig. 7. Human visual latency: data from Mansfield (1973a, Fig. 3 and Table 2) replotted to show how the effect of stimulus area can be accounted for by a single underlying model function. The triangles refer to data obtained with a 0.05-deg foveal spot, the circles and squares to data obtained with 2- and 4-deg foveal spots, large enough for maximal spatial summation. The stimulus, a white 30-ms flash, represents a step of light at all except the very lowest stimulus fluxes. A, the reducible latencies fitted with power functions \( \log(L - L_a) vs \log(L_a) \). Log \( I = 0 \) corresponds to 100 cd/m². \( \beta(0.05 \text{ deg}) = 0.49, \beta(2 \text{ deg}) = 0.31 \). To avoid clutter, the 4-deg line (\( \beta = 0.32 \)) has not been drawn. B, the full latencies as functions of log relative stimulus intensity \( L \ vs \ log(L_a) \). C, corrected reducible latencies as functions of estimated log relative flux \( (L - d vs \log(\phi/\phi_T)) \). The curve is the model “step” latency function for \( n = 4 \) and \( \tau = 100 \) ms. See text for details.

A good test of the model’s explanatory power whether such effects can be accounted for by the single generating mechanism proposed.

Area and duration effects have been most extensively documented by Mansfield (1973b), and Fig. 8 is a replot of data from his Figs. 3 and 4. As in Fig. 7, triangles, circles, and squares mark data obtained with 0.05-, 2-, and 4-deg foveal stimuli, respectively. Panel A shows step functions (“steady state” in Mansfield’s terminology), and panel B shows flash functions (“transient state”). The data were originally fitted with power functions having the following exponents: \( \beta(\text{step}) = 0.52 \) (0.05 deg), 0.34 (2 deg), and 0.34 (4 deg); \( \beta(\text{flash}) = 0.94 \) (0.05 deg), 0.50 (2 deg), and 0.53 (4 deg). In Fig. 8, the data has been flux corrected, and also shifted vertically on the log brightness axis: 1) the flux correction is exactly the same as for the latency data in Fig. 7; and 2) the 2- and 4-deg data have been shifted upwards on the log brightness axis (by 0.5 log units in A and 0.9 log units in B in relation to the 0.05-deg data). These vertical displacements are completely arbitrary and done only to make evident how the different slopes of the functions may arise from different segments of a common underlying function; they entail no claim that the perceived brightnesses of the large and the small stimuli would really have been related as shown, had they been cross correlated.

The continuous lines are model curves, with the parameter \( n = 4 \) obtained from the latency fit in Fig. 7. They do indeed predict the main effects of duration and area: the flash curve is steeper than the corresponding step curve, and the small-spot functions, covering lower flux ranges, are steeper than the large-spot functions. The fits may be less than perfect, but it has to be remembered that the experiments involved a free-floating rationale and extensive averaging across sessions and subjects, where threshold (unknown) must have fluctuated considerably (cf. Marks & Stevens, 1966).

Discussion

From saturating receptors to nonsaturating brightness functions

Only the photoreceptors have access to light, all subsequent stages in the visual system have to make do with signals from the receptors or more proximal neurons. Thus, given present knowledge of rod and cone photoresponses, it seems worthwhile to use them rather than light pulses as such input to the “black box” determining latency and brightness. On the simple assumption that a device summatting linearly in space sets a criterion amplitude and performs a quasi-differentiation at the moment the criterion is reached, a large body of experimental data is explained by the waveform of those responses. Interestingly, the compression of the physical intensity range into a narrower response range follows from the fact that receptor responses are “read” at earlier times the higher the intensity of
stimulation, with no necessary involvement of saturating nonlinearities. Accordingly, although receptor amplitudes saturate after some 3 log units, the predicted functions are nonsaturating, as are human brightness functions over at least 6 log units.

The impulse frequency of frog ganglion cells, as opposed to human brightness sensation, does saturate some 3 log units above threshold. With large stimuli, this can occur when each rod gets less than 10 isomerizations per second, i.e. at intensities 1-2 log units below rod saturation (cf. Hemilä, 1977). As such, it is not surprising that a frog ganglion cell at 11°C is unable to discharge at a very high rate. It points to a problem, though, because no ganglion cells with a 6-log-unit dynamic range have been found in any vertebrate retina (for discussion, see Barlow & Verrillo, 1976). Nevertheless, in monkey LGN there are cells whose initial impulse frequency grows monotonically with intensity over at least 5 log units (Marrocco, 1975; Barlow et al., 1978).

**Comparison with parameters obtained in recordings from single photoreceptors**

The fitting of the model to ganglion cell or psychophysical latencies yields absolute kinetic parameter values $\tau$ and $n$. Save for an amplitude parameter, these uniquely define the shape of the underlying receptor response. Thus, it is a crucial question whether they are consistent with the kinetics of relevant photoreceptor responses recorded under comparable conditions.

**The time constant $\tau$**

$\tau$ characterizes the absolute time scale of responses. Through eqn. (6), $\tau$ can be translated into the time-to-peak ($t_p$) of the receptor's dim-flash response, allowing easy comparison with recorded receptor responses. (It does not matter much if the flash response of the receptor would, in fact, be better fitted by some other linear multistage filter model than that which assumes "independent activation," because the differences are mainly connected with the falling phase.) The mean values of $\tau$ here obtained for frog ganglion cells would indicate times-to-peak from 1.5 s (for the responses with $n = 6$) to 2.5 s (for $n = 4$) at 11°C. Recordings of the photocurrent from single rods of the closely related species *Bufo marinus* have given $t_p = 1.3$ s at 20°C (Baylor et al., 1979, $n = 4$). With a $Q_{10}$ around 2 (Lamb, 1984; Donner et al., 1988), this gives $t_p = 2.4$ s at 11°C. Lamb's (1984) recordings from the same species at about 11°C indicates $t_p = 2$ s (in bicarbonate solution).

The time-to-peak of intracellularly recorded voltage responses shortens sharply as the flash intensity is raised, curtailed by mechanisms which do not affect the early rise of the response and are not foreseen by the independent activation model (see Schwartz, 1976; Cervetto et al., 1977; Detwiler et al., 1980). Thus, only responses to one or a few isomerizations per rod are useful for the present comparison. In such voltage recordings from sufficiently sensitive *Bufo marinus* rods (Reuter et al., 1986; Donner et al., 1989 and unpublished observations), the mean $t_p$ was about 1.6 s (range 0.8–2.7 s in seven cells) at 20°C, suggesting 3 s (1.5–5 s) at 11°C. Intracellular recordings
at 11°C from rods of the species used here gave dim-flash times-to-peak from 1.7–4 s (K. Donner, S. Hemila & A. Koskelainen, four unpublished experiments). These comparisons show that the rod kinetics inferred from model fits to dark-adapted ganglion cells is well consistent with that found in recordings of either the membrane voltage or the outer segment current of single rods.

The same conclusion applies to the model fit to human reaction times (Fig. 7), which required $\tau$ around 100 ms. This $\tau$ must predominantly reflect rod kinetics, because white stimuli were presented in darkness to a dark-adapted observer. Rods will then determine the latencies over a considerable range from threshold upwards, and this range (with the longest latencies) has most weight in the determination of $\tau$. Baylor et al. (1984), fitting current responses from single monkey rods (Macaca fascicularis, 37°C) with the independent activation model, found $\tau$ values 100–130 ms. The present author has also fitted latencies of rod-mediated ganglion cell responses in the retina of Macaca mulatta with the model presented here (37°C; data from Gouras & Link, 1966). The optimal fit was achieved with $\tau = 110-120$ ms, in good agreement with the above values.

The number of steps $n$

Averaged human data, as well as data from some individual human subjects and some frog ganglion cells, were best described by assuming the number of stages $n = 4$ in the receptors. This is in agreement with current recordings from rat, toad, and frog rods (Penn & Hagins, 1972; Baylor et al., 1979; Donner et al., 1988). On the other hand, intracellular voltage recordings from toad and turtle rods and turtle cones, and current recordings from monkey rods, have consistently required $n = 6$ or 7 (Cervetto et al., 1977; Schwartz, 1976; Baylor et al., 1974, 1984). Indeed, $n$ values 6 and 7 were encountered in a substantial fraction of individual human subjects and frog ganglion cells. When judging this variation, it is worth noting that two effects are liable to specifically decrease apparent $n$ values:

1. As soon as there is a temporal spread of the signals from different elements, the earliest rise of a summed response from the ensemble will suggest an $n$ value lower than that of the elementary response. Assume, for example, that the response of a single receptor rises as $a t^b$ (see eqn. (3)), and transmission delays ($D \pm \delta_t$) are normally distributed around the mean $D$ with standard deviation 10 ms (cf. Baylor & Fettiplace, 1977). Then the summed response will rise proportionally to $t^2$ for some 25 ms, proportionally to $t^4$ for the next 20 ms, and only after that reach the "proper" dependence on $t^6$ (as seen from a serial expansion of the expression $a(t + \delta_t)^6$ with $\delta_t$ normally distributed around 0).

2. In reality, brightness must be determined by the rise of the receptor response over a finite interval, not by derivatives. This would make the model curves steeper in the high-intensity end, the more so the longer the interval (subject to individual variation). By comparison, the derivative-based templates used here will tend to underestimate $n$.

Thus it may be that, in general, the highest $n$ estimates reflect receptor responses more faithfully than lower ones. If so, $n = 6–8$ is suggested not only by frog ganglion cells and human psychophysics, but also by the pupil reflex and LGN cells of the macaque (Barlow et al., 1978; a power-function analysis such as used here applied to their LGN latency data indicates $\beta = -0.16$). Still, it must be remembered that the particular formulation of the model given here only represents a simplest variant: intervening differentiations and/or integrations of the receptor signal would leave the predictions essentially unchanged, but could alter the apparent $n$.

Earlier related studies

The proposed model is based on receptor responses simulated by a cascade of linear low-pass filters as first suggested by Fuortes and Hodgkin (1964) and subsequently adopted by a majority of recent studies. Other authors have previously pointed out the possibility of relating psychophysical latency (Alpern, 1968) or brightness (Sperling & Sondhi, 1968; Marks, 1972) functions to this type of model, but these authors were not concerned with the physiological basis, and the correlations were left on a general level.

Mansfield and Daugman (1978), on the other hand, directly showed that the times to a criterion amplitude of the mass receptor response in toad could approximate cube-root functions of intensity, much like psychophysical latencies. From this they argued that the dominant mechanism controlling the intensity-dependent portion of visual latency lies in the photoreceptors. However, they did not critically examine the limitations of the power-function approach, and were therefore led to conclude that linear-filter cascade models for photoreceptor responses were incompatible with their results.

In a recent study of human reaction time (as a function of contrast), Burkhardt et al. (1987), while fitting their data with power functions, noted that the Baylor–Hodgkin–Lamb (1974) model could in principle account for the results. They pointed out that "... a photoreceptor that is linear at the very first stage of transduction can ... generate a nonlinear signal at its output that displays an approximate power law" and concluded that their quantitative findings on human reaction time (in light-adapted conditions) might be largely shaped by cones. Interestingly, they showed that negative contrasts generate similar latency functions as positive ones (cf. also Burkhardt et al., 1984; Burkhardt & Gottesman, 1987). This is to be expected whenever the receptor's OFF response is the (subsequently rectified) mirror-image of the ON response (as in eqn. (42) of Baylor et al. (1974)), i.e. when saturation effects can be neglected.

Acknowledgments

I am grateful to Dr. Karl Johan Donner for the computer program for fitting power functions, and to Ms. Annika Luther, M.Sc., for drawing the figures. I also wish to thank Drs. Simo Hemilä and Tom Reuter for their helpful comments on the manuscript. This work was supported by the Academy of Finland (Grant 01/455).

References


Latency and brightness functions


