THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 9, November, 1973
Printed in U.S.A.

Model Approximations to Visual Spatio-Temporal Sine-Wave Threshold Data

By Z. L. BUDRIKIS

(Manuscript received May 8, 1973)

Experimental data on visual spatio-temporal sine-wave thresholds obtained by Robson and Kelly are considered. In seeking model approximations to the data it is assumed that the subject's visual threshold to modulation at different spatial and temporal frequencies gives the image of his filter function to within a multiplicative constant. It is further assumed that the data can be approximated by a system with a spatially uniform, isotropic, and temporally invariant response which consists of the difference between an excitatory and an inhibitory term, and that each term is separable into a product of a spatial and a temporal function.

I. INTRODUCTION

Tests of vision with sine-wave flicker go back at least fifty years to H. E. Ives.[1] He determined flicker fusion frequencies with a number of wave shapes, including sinusoids. Spatial sinusoid test stimuli are more recent. The first to use them was probably Schade[2] in the fifties. Soon after that Kelly[3] suggested a stimulus which would simultaneously test the spatial and the temporal sine-wave response of vision. Such tests were implemented by Robson,[4] Kelly,[5,6] and others.

The special interest in the sine wave as a test stimulus stems from the ease with which one can extrapolate from its results. Provided a system is linear and time-invariant, Fourier analysis can be used to predict the system response to any input from its response to sinusoidal inputs. However, the visual system is neither linear nor time-invariant. Nevertheless, given a sufficiently constant adaptation state and input variations that result in small output variations,† linear theory can be used.

† It is often incorrectly stated or implied that, for linearity, the input needs to be small. But consider the situation where a flickering light appears fused visually. The input may then swing between zero and many times the average luminance, yet the behavior is linear.

Our interest in the visual system is related to visual communications. When visual messages are transmitted digitally then there are potentially very many different ways—some more advantageous than others—in which the messages might be coded and still give acceptable fidelity at the receiver. Clearly, it would be good if the likely subjective effect of given quantizing procedures could already be predicted at the computer simulation stage without involving repeated subjective tests. Such predictions will probably be possible soon.[7] However, there is still need for complete specifications both of the linear behavior of the visual system and of the nonlinear effects of background masking. We will concern ourselves here only with the linear characteristics. To this end we will examine several alternative mathematical models to see whether they could be used to represent published experimental data on spatio-temporal sine-wave thresholds.

The data that we will use were reported by Kelly[6] and Robson.[4] In both cases threshold values of m were determined in a target described by

$$L = L_o(1 + m \cos 2\pi u_o x \cdot \cos 2\pi f_o t) , \tag{1}$$

where L_o is the average luminance, u_o the spatial frequency, and f_o the temporal frequency.

Kelly's measurements† were made at four different values of L_o. The entire target area, a circular 7-degree CRT face, filled with the flickering grating, was viewed monocularly through a 2.3-mm artificial pupil. Robson made all measurements at a single L_o value. The target had a 2.5-degree × 2.5-degree grating in the center of a 10-degree × 10-degree screen which had a luminance equal to L_o, and it was viewed binocularly without artificial pupils.

In both cases the subject's threshold was measured by the method of adjustment. The subject judged whether he could see the signal or not. He did not attempt to distinguish between seeing flicker and seeing the bar pattern. During each session of Kelly's experiment, the subject made 5 settings at each of 12 frequencies, with the 60 presentations given to him at random. Robson made his measurements in orderly sequences. Their results are shown as log-log plots of $(1/m)$ against frequency in Figs. 1–5.

Kelly's measurements obtained for vision with an artificial pupil are converted to equivalent luminances viewed through a natural pupil. To calculate the equivalent luminance one needs to take into account changes in the size of the natural pupil and the Stiles-Crawsford effect. From data tabulated by LeGrand[8] it can be inferred that,

† D. H. Kelly kindly supplied a listing of his measurements and standard deviations.

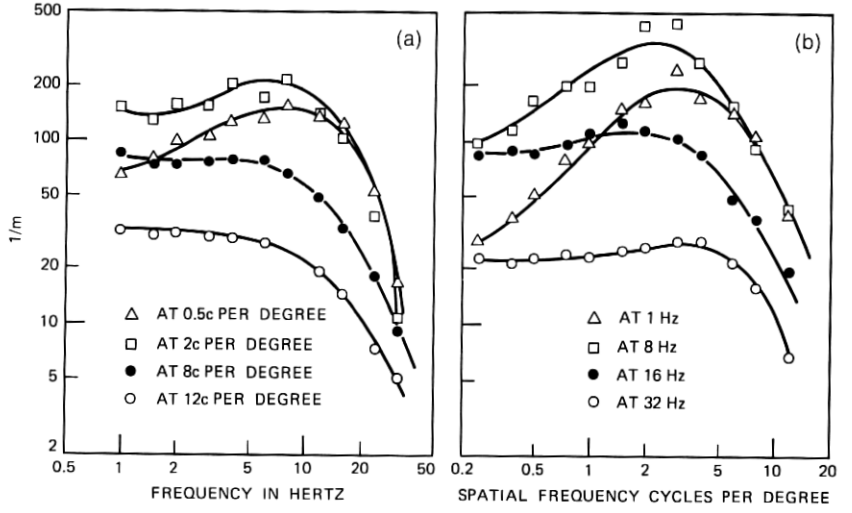


Fig. 1—Kelly's data at 62.8 mL. (a) Temporal frequency response. (b) Spatial frequency response.

given an illuminance I, in trolands, the corresponding luminance L in mL is

$$L = 1.142 \times 10^{-2} I^{1.223}; \qquad 10 < I < 2000 \text{ td.} \tag{2}$$

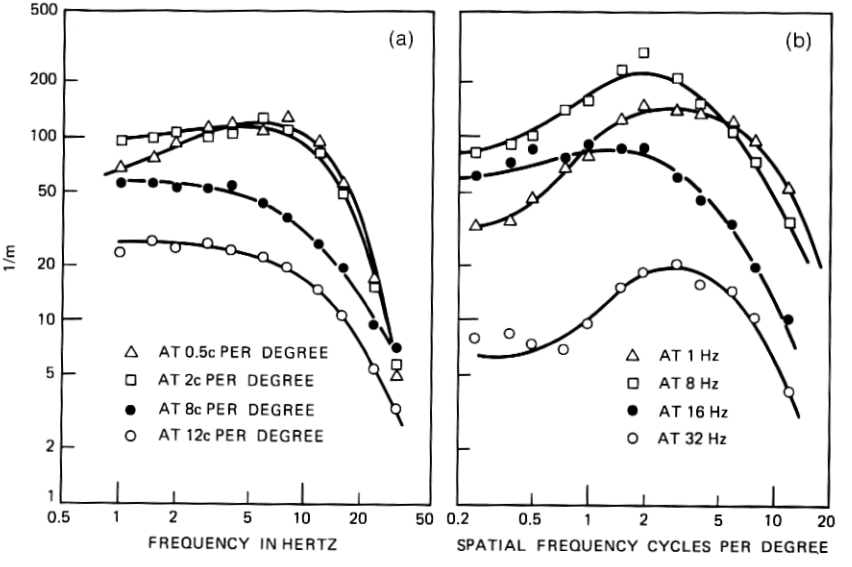


Fig. 2—Kelly's data at 15.2 mL. (a) Temporal frequency response. (b) Spatial frequency response.

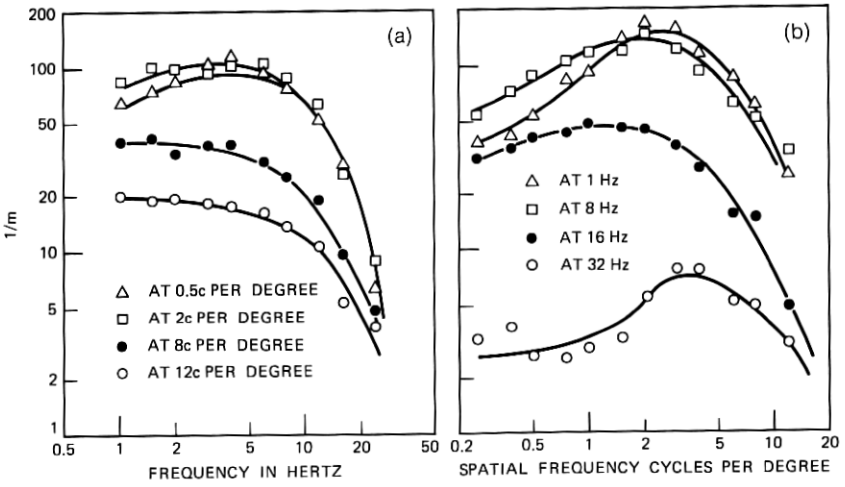


Fig. 3—Kelly's data at 3.7 mL. (a) Temporal frequency response. (b) Spatial frequency response.

We will consider six different, though similar, mathematical models as possible candidates for representing the data of Figs. 1–5. There is a similarity between the models in that: (*i*) they all consist of an algebraic difference between an excitatory and inhibitory term, (*ii*) these terms are in all models separable functions of spatial and tem-

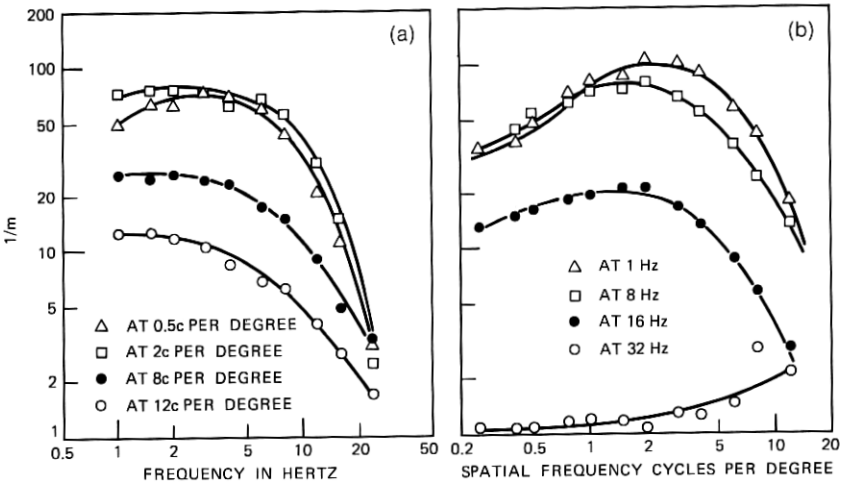


Fig. 4—Kelly's data at 0.91 mL. (a) Temporal frequency response. (b) Spatial frequency response.

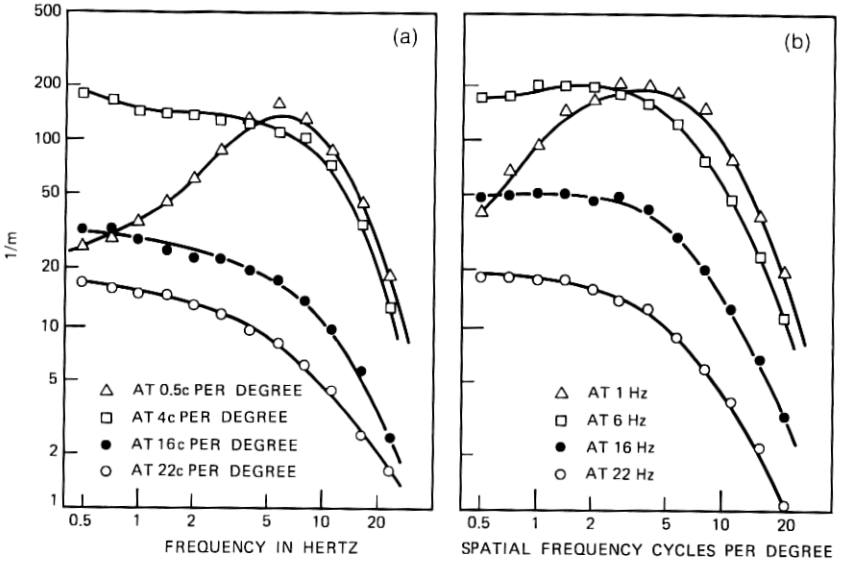


Fig. 5—Robson's data at 6.3 mL. (a) Temporal frequency response. (b) Spatial frequency response.

poral frequencies, and (*iii*) in each model there will be six undetermined parameters. We find values for the parameters by digitally searching for the smallest weighted mean-square deviation of experimental points from the models. The fit of none of the models is compellingly good, but in several cases the degree of fit is useful. We find the best all-round fit with a model with diffusion-like temporal response of excitation, a Gaussian function for the temporal response of inhibition, and Cauchy functions for the spatial response. From the point of view of economy in computer simulation, a model with simple exponential time responses and Gaussian spatial responses would be preferable. However, the mean-square departure from the model is somewhat larger than the best.

II. THE MODELS

2.1 *The Framework*

By the nature of things, the retinal image is a somewhat blurred version of the light distribution in object space. Over isoplanatic patches,[9] or areas A which are large compared to the size of a blurred point and small compared to inhomogeneities of the image-forming properties of the eye, we can model the formation of the image by a

convolution integral with a fixed point-spread function:

$$I(x, y) = \iint_A W(x - \xi, y - \eta)L(\xi, \eta)d\xi\, d\eta\,, \tag{3}$$

where W is the point spread function and L and I are the object and image distributions.

There is virtually no time lag in forming the retinal image. If L were switched on at some instant of time then the retinal image $I(x, y)$ would be formed at that instant.

The object-space to image-plane spatial frequency response of the given isoplanatic patch, or its modulation transfer function, is the Fourier transform of W:

$$H(u, v) = \int\int W(x, y)e^{-2\pi j(ux+vy)}dx\, dy\,, \tag{4}$$

where u and v are the spatial frequencies in the x and the y directions. Point spread in the space domain becomes filtering when transformed into the frequency domain. The point spread function W is necessarily positive and, with a normal pupillary aperture, has a maximum at the center and decreases monotonically.[10] Consequently $H\ (u, v)$ is a low-pass function.

It is natural to think of perception being based on an "image" at some deeper location beyond the retina. This "image" is physiologically mediated and must suffer appreciable time lags. Hence, the response at the deeper location will be time-dependent. There will also be further spatial filtering as a result of lateral physiological interactions.[11]

Say we designate the resulting point response function by $R(x, y, t)$ and the internal "image" distribution by $C(x, y, t)$. At least for a restricted class of object-space luminance functions, $L(x, y, t)$, C can be obtained by superposition, so that

$$C(x, y, t) = \int_A \int\int_o^{+\infty} R(x - \xi, y - \eta, t - \tau)L(\xi, \eta, \tau)d\tau\, d\eta\, d\xi. \tag{5}$$

The three-dimensional Fourier transform of R is the spatio-temporal frequency response function

$$S(u, v, f) = \int\int\int R(x, y, t)e^{-2\pi j(ux+xy+ft)}dx\, dy\, dt. \tag{6}$$

The integration is over all x, y, and t. f is the temporal frequency.

We may assume that the response function R is even in x and y, i.e., $R(x, y, t) = R(-x, y, t) = R(x, -y, t) = R(-x, -y, t)$. This

means that S is even in u and v, i.e., $S(u, v, f) = S(-u, v, f) = S(-u, -v, f)$. No symmetry can be assumed for R in t, and hence, for S in f. Indeed, $R(x, y, t) = 0$ for $t \leqq 0$, and hence, $S(u, v, -f) \neq S(u, v, f)$.

If the input to the system is the L of eq. (1), then the internal "image" is

$$C(x, y, t) = S(0, 0, 0)L_o + |S(u_o, 0, f_o)|L_o m \cos(2\pi u_o x) \times \cos(2\pi f_o t + \phi), \quad (7)$$

where

$$|S(u_o, 0, f_o)| = \{S(u_o, 0, f_o) \cdot S^*(u_o, 0, f_o)\}^{\frac{1}{2}}$$

and

$$\phi = \tan^{-1}\{\mathrm{Im}[S(u_o, 0, f_o)]/\mathrm{Re}[S(u_o, 0, f_o)]\},$$

The * designates the complex conjugate and Im and Re the imaginary and real parts.

Now we ask: What size must m be before the flickering grating is seen with a given level of certainty? We assume thresholds correspond to fixed differences, i.e., the flickering grating is seen with probability p if

$$|S(u, 0, f)|L_o m = T(p), \quad (8)$$

where T is a monotonically increasing function of p, but is independent of all other variables. We may assume that subjects adjusted m so that it always resulted in the same probability of seeing. Therefore, the values of $1/m$, as plotted in Figs. 1–5, are regarded as experimental determinations of $|S(u, 0, f)|$ [to within the multiplier $T(p)/L_o$ which is a constant when the criterion T and the average luminance L_o are fixed].

If the visual system were truly linear it would have the same response functions irrespective of luminance level L_o. But all evidence, including that contained in Figs. 1–4, shows that the system adapts. It does so somewhat ponderously, much faster with rising L_o than in reverse, but still quite effectively, changing gain, spatial spread, and temporal lag. There is just one aspect of S which Kelly[12] found unchanging over more than four decades of luminance, L_o. In large-area flicker threshold determinations, using an artificial pupil, he found that at different L_o values plots of $(1/mL_o)$ approached a common asymptote for large values of f. However, in other parts of the functional domain, different S functions hold for different adaptation luminances.[6]

In searching for suitable mathematical expressions for R or S it would be convenient if these functions were isotropic, and even more

so, if they were also separable into spatial and temporal factors. Isotropism would mean that the space variables x and y would reduce to a single distance ρ and the frequencies u and v to a direction-independent spatial frequency ν. Then

$$S(\nu, f) = S(u = \nu, 0, f) = S(0, v = \nu, f)$$
$$= \int_0^\infty \left[\int_0^\infty R(\rho, t) 2\pi\rho J_o(2\pi\rho\nu) d\rho \right] e^{-j2\pi ft} dt , \tag{9}$$

where J_o is the Bessel function of order zero.

Man's vision is not isotropic. It is astigmatic, having better resolution in the horizontal and vertical directions than at other angles. But, to a first order of approximation, we may assume isotropism.

Separability of R would mean that we could write it as

$$R(\rho, t) = U(\rho)V(t) \tag{10}$$

and then S would also be separable:

$$S(\nu, f) = G(\nu)H(f) , \tag{11}$$

where

$$G(\nu) = \int_0^\infty U(\rho) 2\pi\rho J_o(2\pi\rho\nu) d\rho \tag{12}$$

and

$$H(f) = \int_0^\infty V(t) e^{-j2\pi ft} dt . \tag{13}$$

Moreover, because $U(\rho)$ is symmetrical, and hence $G(\nu)$ is a real-valued function, it would follow that

$$|S(\nu, f)| = G(\nu)|H(f)| . \tag{14}$$

However, even a superficial look at the families of experimental curves in Figs. 1–5 will convince one that $|S(\nu, f)|$ is not separable. If it were, then curves of $|S(\nu, f)|$, as functions of f at different values of ν, would differ from each other only by constant multipliers. Plotted against a logarithmic ordinate this would result in fixed vertical shifts. The same result would hold for plots of $|S(\nu, f)|$ versus ν at different values of f. But neither of these outcomes are found to be true. This is particularly evident when looking at Figs. 5a and b. The curves at high values of ν or f are low-pass in shape, while for low values of the parameters they are bandpass. Figure 6 shows a linearly scaled perspective view of a surface[13] to which the measured values of Fig. 5 approximate. Measurements apply only to positive frequencies, while

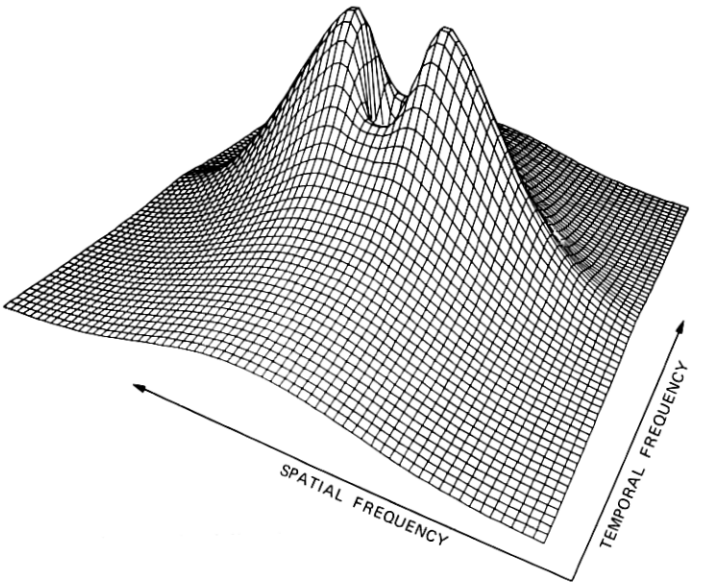


Fig. 6—Perspective view of spatio-temporal frequency response.

the surface has been drawn over all four quadrants making use of symmetry. It suggests a volcano with a deep central crater.

It is customary[7] to think of the response as being brought about by an interplay of excitation and inhibition, with inhibition responsible for the crater. Looked at in this way the measurements suggest, at least to a first approximation, that excitatory and inhibitory responses in themselves may be separable and that the effects of inhibition simply subtract from the effects of excitation. These assumptions will be made. The response functions can then be formally broken down:

$$\begin{aligned} R(\rho, t) &= R_e(\rho, t) - R_i(\rho, t) \\ &= U_e(\rho)V_e(t) - U_i(\rho)V_i(t) \end{aligned} \tag{15}$$

$$\begin{aligned} S(\nu, f) &= S_e(\nu, f) - S_i(\nu, f) \\ &= G_e(\nu)H_e(f) - G_i(\nu)H_i(f) \end{aligned} \tag{16}$$

with

$$G_e(\nu) = \int_0^\infty U_e(\rho)2\pi\rho J_o(2\pi\rho\nu)d\rho,$$

$$H_e(f) = \int_0^\infty V_e(t)e^{-j2\pi ft}dt,$$

and similarly for the inhibitory functions.

2.2 *Choice of Functions*

To satisfy physical considerations, all the component functions should be low-pass in character. Of the immense number of possibilities we consider just several. The Gaussian function comes readily to mind particularly for spatial spreads.

If

$$U_e(\rho) = \frac{1}{2\pi a^2} e^{-\rho^2/2a^2}, \tag{17}$$

then of course the Fourier transform is also Gaussian:

$$G_e(\nu) = e^{-2\pi^2 a^2 \nu^2}. \tag{18}$$

The function has another property which can be especially useful in computations, namely that as a function of two variables x and y, i.e., $\rho^2 = x^2 + y^2$, it is separable:

$$U_e(x, y) = \left(\frac{1}{\sqrt{2\pi a^2}} e^{-x^2/2a^2}\right)\left(\frac{1}{\sqrt{2\pi a^2}} e^{-y^2/2a^2}\right). \tag{19}$$

None of the other functions of interest to us has this property.

Another possible candidate for point spreads is the exponential

$$U_e(\rho) = e^{-2\pi b\rho}; \qquad p > 0 \tag{20}$$

and then

$$G_e(\nu) = \frac{b}{2\pi(b^2 + \nu^2)^{\frac{3}{2}}}. \tag{21}$$

At high values of ν, $\nu \gg b$, the function decreases as $(1/\nu)^3$ which corresponds to a fall-off of 18 dB/octave.

On the other hand, if the spatial frequency response function were an exponential then there would be no straight-line asymptote on a log-log plot, but rather a response which would be

$$U_e(\rho) = \frac{c}{2\pi(c^2 + \rho^2)^{\frac{3}{2}}} \tag{22}$$

with

$$G_e(\nu) = e^{-2\pi c|\nu|}. \tag{23}$$

This is often called the Cauchy response.

Since the temporal frequency responses are similar to the spatial frequency responses similar functions can be used to model these. The important differences are that the function $V(t)$ is one-sided and that eq. (13), instead of (12), is used to obtain the Fourier transform.

The Gaussian function can be used in an approximate way by shifting it a distance t_o to the right along t and deleting it leftward of $t = 0$:

$$V_e(t) = \frac{1}{\sqrt{2\pi}\tau} e^{-(t-t_o)^2/2\tau^2}; \qquad t \geqq 0$$
$$= 0; \qquad t < 0. \tag{24}$$

When t_o/τ is greater than three, say, then there is negligible error in assuming that $V_e(t)$ is the Gaussian function for all t, $t < 0$ included. Then

$$H_e(f) = e^{-2\pi^2\tau^2 f^2 - j2\pi f t_o}. \tag{25}$$

In computer simulation a simple exponential time response, often known as the Poissonian, would be the easiest because it can be effected by recursion. That function and its transform are

$$V_e(t) = (1/\tau_1)e^{-t/\tau_1}; \qquad t \geqq 0 \tag{26}$$

$$H_e(f) = \frac{1}{1 + j2\pi f\tau_1}. \tag{27}$$

A function in which there is theoretical interest[12,14] is one that occurs in diffusion processes. Kelly[12] found that the high-frequency asymptote for large-area flicker responses could be fitted well with a frequency function which one would find in diffusion that had no losses in the diffusing substance, namely with

$$H_e(f) = C_1 e^{(-|2\pi f\tau|^{\frac{1}{2}})}. \tag{28}$$

If the Laplace transform is taken as

$$H_e(s) = C_1 e^{-(2s\tau)^{\frac{1}{2}}}, \tag{29}$$

then the time function is [12,15]

$$V_e(t) = \frac{\tau^{\frac{1}{2}} e^{-\tau/2t}}{(2\pi)^{\frac{1}{2}} t^{\frac{3}{2}}}; \qquad t \geqq 0. \tag{30}$$

The six models which were compared with the experimental data are:

(*i*) Gaussian temporal/Cauchy spatial (G/C)

$$|S(\nu, f)| = Ae^{-2\pi^2 f^2\tau_1^2}(e^{-\nu\sigma_e} - ke^{-2\pi^2 f^2\tau_2^2} e^{-\nu\sigma_i}), \tag{31}$$

(*ii*) Poissonian temporal/Cauchy spatial (P/C)

$$|S(\nu, f)| = \frac{A\{[e^{-\nu\sigma_e}(1 + 4\pi^2 f^2\tau_2^2) - ke^{-\nu\sigma_i}]^2 + (2\pi f\tau_2 ke^{-\nu\sigma_i})^2\}^{\frac{1}{2}}}{(1 + 4\pi^2 f^2\tau_1^2)(1 + 4\pi^2 f^2\tau_2^2)^{\frac{1}{2}}}, \tag{32}$$

(*iii*) Diffusion-Gaussian temporal/Cauchy spatial (*D-G/C*)

$$|S(\nu, f)| = A(e^{-(f\tau_1)^2}e^{-\nu\sigma_e} - ke^{-2\pi^2 f^2\tau_i^2}e^{-\nu\sigma_i}), \tag{33}$$

and three further models in which the Gaussian is substituted for the Cauchy response giving G/G, P/G, and D-G/G.

Note that in four of the models, G/C, P/C, G/G, and P/G, one time-lag stage is common to excitation and inhibition (Fig. 7a). The remaining two models, involving diffusion, have distinct paths for the two effects (Fig. 7b).

Each of the models differs from the others in its exact functional shapes but they are all similar in their form. Figure 8 illustrates the evolution of the point spread as given by the P/G model. The point spread function is shown at the instant of occurrence of the point impulse and at two subsequent time instants thereafter. In this, as in all the other models, the excitatory effect is confined to a smaller region and has a faster time course than the inhibitory effect.

III. SELECTION OF PARAMETERS

Each of the models chosen for comparison with the experimental data has six undetermined parameters: the gain A, time constants τ_1 and τ_2, space constants σ_e and σ_i, and the per unit inhibition k. The

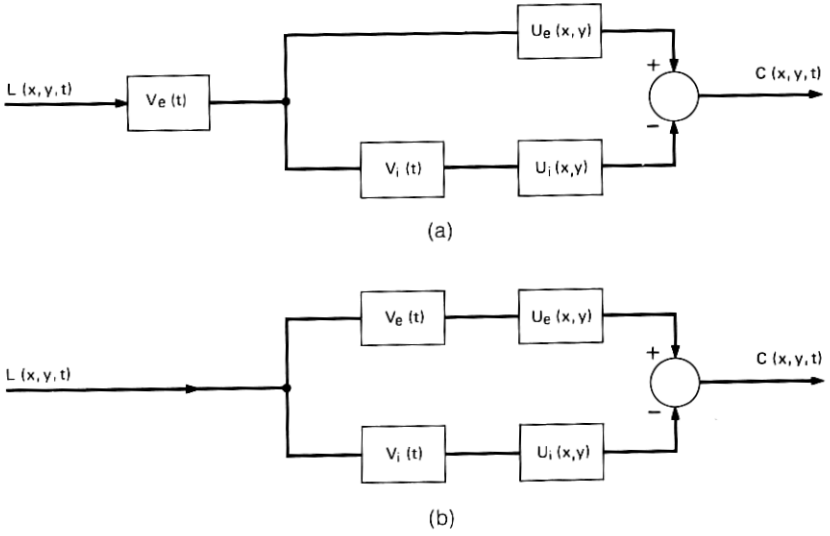


Fig. 7—System block diagram: (a) for P/G, G/G, P/C, and G/C models; (b) for D–G/G and D–G/C models.

parameters have to be given values to produce as close a fit as possible between model and data.

The following performance index may be used as an appropriate measure for the closeness of fit:

$$P = \sum_{i=1}^{N} \{[m_i - 1/|S(\nu_i, f_i)|]/\epsilon_i\}^2, \tag{34}$$

where m_i is the measured threshold modulation at the spatial frequency ν_i and temporal frequency f_i and ϵ_i is the estimated (standard) error of that measurement. The summation is over all N points measured at a given luminance L_o.

This will be called the aggregate-square fractional error, or ASFE, index. The ASFE index is perhaps the most defendable in light of the experimental procedure. However, if the aim is to obtain the best representation of data plotted as $(1/m)$ along a logarithmic scale (Figs. 1–5), then a better index is

$$P = \sum_{i} \{\log [|S(\nu_i, f_i)|/m_i]\}^2, \tag{35}$$

which can be called the aggregate-square log error, or ASLE, index.

Irrespective of index, the array of six parameter values can be looked upon as a vector $\mathbf{T}$ and the performance index as a real-valued function of it. Our object is then to find that location $\mathbf{T}_m$ in six-space at which $P(\mathbf{T})$ assumes its smallest value. However, there is no way of recognizing a global minimum and it is therefore impractical to insist on finding it. The object is rather to find as good a value for $\mathbf{T}$ as possible, while keeping computer expenditures within reasonable bounds.

Of the many possible parameter search routines we tried a gradient-dependent algorithm, random search, and a combination of the two. Random search proved the more successful, almost as good on its own as in combination with gradient techniques.

The gradient in question is

$$\nabla P = \sum_{n=1}^{6} \frac{\partial P}{\partial T_n} \mathbf{a}_n, \tag{36}$$

where $\mathbf{a}_n$ is the unit vector along the nth coordinate axis and T_n is the scalar $(\mathbf{T} \cdot \mathbf{a}_n)$. The components of the gradient were evaluated in one of two ways:

(*i*) approximate differentiation:

$$\frac{\partial P}{\partial T_n} \doteq \frac{P(\mathbf{T} + \Delta T_n \mathbf{a}_n) - P(\mathbf{T})}{\Delta T_n}; \tag{37}$$

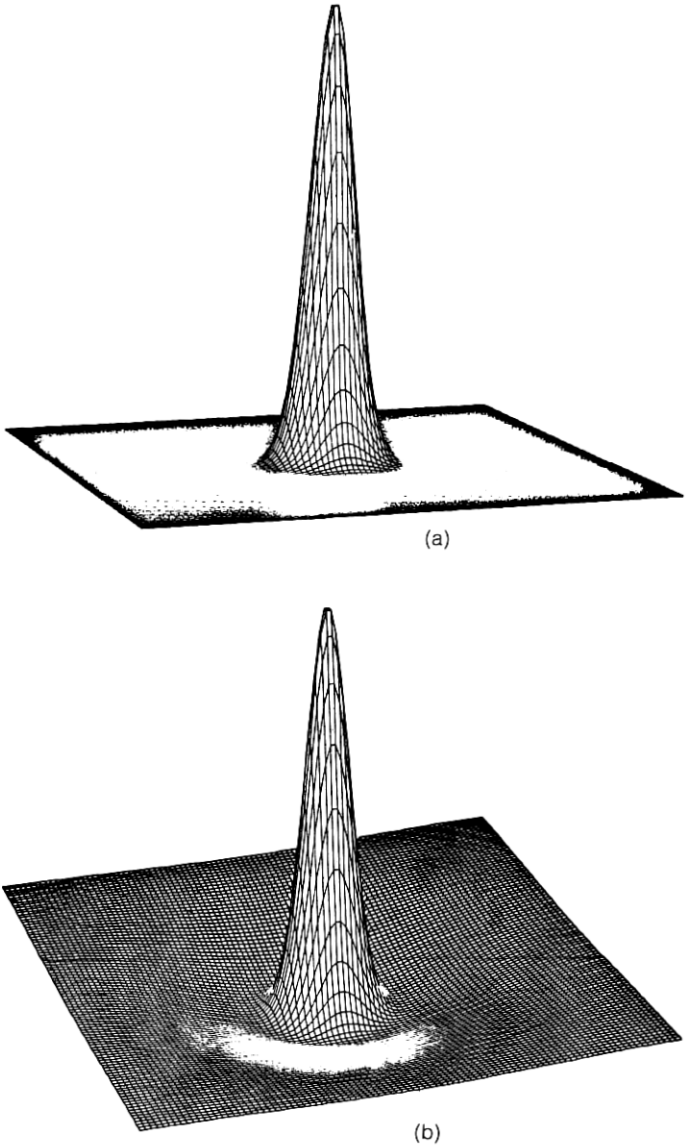
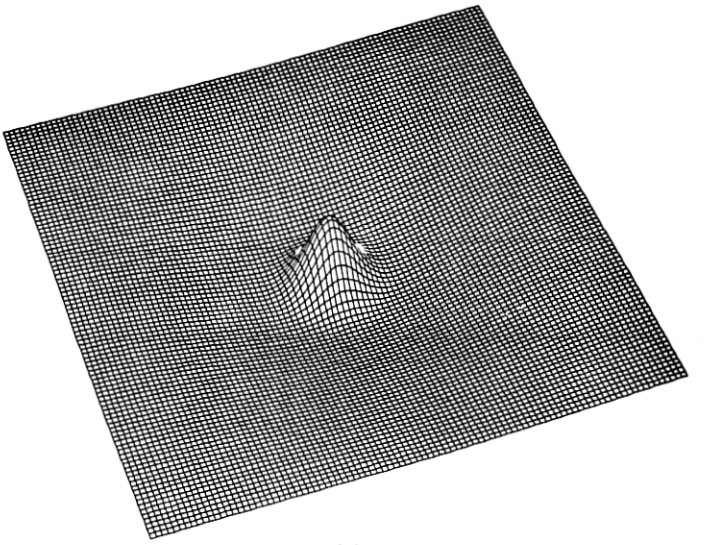


Fig. 8—Evolution of the point spread function in the Poissonian/Gaussian model. Inhibitory effect has been exaggerated. (a) at $t = 0$, (b) at $t = 45$ ms, (c) at $t = 150$ ms.



(c)

Fig. 8 (continued).

(*ii*) evaluation of exact expressions which, given (34), are

$$\frac{\partial P}{\partial T_n} = \sum_i 2\{[m_i - 1/|S(\nu_i, f_i)|]/\epsilon_i\} \times [1/\epsilon_i |S(\nu_i, f_i)|^2] \frac{\partial |S(\nu_i, f_i)|}{\partial T_n}. \quad (38)$$

Using (37), care had to be taken in choosing the size of ΔT_n.

Given the gradient at the vector location $\mathbf{T}_j$, the next location with a lower value of P should be at

$$\mathbf{T}_{j+1} = \mathbf{T}_j - K\nabla P|_{\mathbf{T}=\mathbf{T}_j}. \quad (39)$$

This will prove to be so, provided K is small enough. Improved convergence rates are possible by making K variable,[16] increasing its value with repeated improvements in P, and decreasing it with failures. The next location to be tested is then not given by (38) but by

$$\mathbf{T}_{j+1} = \mathbf{T}_b - K_j\nabla P|_{\mathbf{T}=\mathbf{T}_b}, \quad (40)$$

where $\mathbf{T}_b$ is the location at which the last lowest value of P was calculated and K_j has been determined from a starting value K_o by multiplication with either $\alpha(0 < \alpha < 1)$ or $\gamma(1 < \gamma)$, depending on outcomes of the j iterations thus far.

A difficulty with gradient-dependent search is that it may end at a local minimum which is far above the global, and that often proved to be so. A way around this is to alternate between the gradient-dependent search mode and random search. In random search the next location to be tested would be

$$\mathbf{T}_{j+1} = \mathbf{T}_b + K \sum_{n=1}^{6} G_n R_j(n) \mathbf{a}_n, \tag{41}$$

where $\mathbf{T}_b$ is again the last best location, K is a constant that scales the size of the search volume, G_n's are further scaling factors designed to make the search about equally sensitive along the different coordinates, and $R_j(n)$ is a Gaussian variate obtained from a (pseudo)-random number routine taking a fresh value for each component and each iteration.

Typically, a computational cycle would consist of gradient-dependent search to within a convergence test specification, taking some 20 to 100 iterations, followed by 100 iterations of random search. The number of cycles depended on progress and could be as many as 50.

Most of the performance improvements were found to come from the random search phases of the computational cycles. For that reason the gradient-dependent phase was dispensed with in many calculations, and then K of (40) became a variable similar to K_j of eq. (39). The calculation was still done in cycles, starting each cycle with a large value of K.

IV. RESULTS

Although there is no guarantee that the performance indexes finally arrived at are the lowest possible, in each case the chances are small that there would be anything substantially lower. Hence, Table I can be taken as a good guide for comparing the effectiveness of the different models in fitting the data. The table gives rms deviations D, which are calculated from P in accordance with

$$D = [P/(N - 6)]^{\frac{1}{2}}. \tag{42}$$

Division is by $(N - 6)$, because the parameters provide six degrees of freedom. For Table I, P was as defined by eq. (34), i.e., the ASFE criterion.

From the last column of Table I it can be seen that the best of the six models is the Diffusion-Gaussian/Cauchy and the worst the Gaussian/Cauchy. The Poissonian/Gaussian is somewhat worse than the average over the group. A comparison of the models by order of

TABLE I—RMS DEVIATIONS

Summary of rms deviations, D, derived from ASFE performance index [eq. (34)]. In computation, experimenter's estimates of experimental errors were used with Kelly's data and assumed errors with Robson's data.

Model \ Luminance (mL)	Kelly's Data				Robson's Data	Mean D for Model
	62.8	15.2	3.7	0.91	6.3	
Poissonian/Gaussian	7.0	4.3	3.9	7.6	3.7	5.3
Poissonian/Cauchy	7.2	4.4	4.0	7.9	3.4	5.4
Gaussian/Gaussian	4.0	4.6	4.7	8.8	2.9	5.0
Gaussian/Cauchy	6.1	6.2	5.6	9.8	4.2	6.4
Diff-Gauss/Gaussian	4.0	3.6	4.3	4.4	3.9	4.0
Diff-Gauss/Cauchy	3.4	3.7	4.5	4.5	2.8	3.8

rank within each set, and then over the sets, shows the two Diffusion-Gaussian models fit best, closely followed by the Poissonian/Gaussian model.

Except with Robson's data, where assumed error values were used, the actual magnitude of D in Table I has significance. With P by eq. (34) being measured relative to experimental errors one would expect with a perfect model fit a D value of unity. $(D - 1)$ is then the increase in relative error due to the model, and D can be thought of as error gain. In this sense all the models, including the best, give only poor fits.

The D-G/C model is shown fitted to Robson's data in Figs. 9a and b. According to Table I this ought to be about the best fit, but obviously is only fair. The same data is fitted by the P/G model in Figs. 10a and b. The P/G model is shown fitted to Kelly's data at 62.8 mL in Figs. 11a and b. According to Table I the P/G model represents nearly the worst fit.

Parameter values for the P/G model are given in Table IIA. These were determined using the relative error criterion. Table IIB gives parameter values for the same model but determined by the log departure criterion. The final mean log departures are shown in the bottom row. There are noticeable differences between the parameter values in Table IIA and Table IIB but, given the rather poor fit between model and data, agreement is good. Consistent trends are apparent in both sets: with decreasing luminance the gain (A) of the system decreases accompanied by a decrease in fractional inhibition

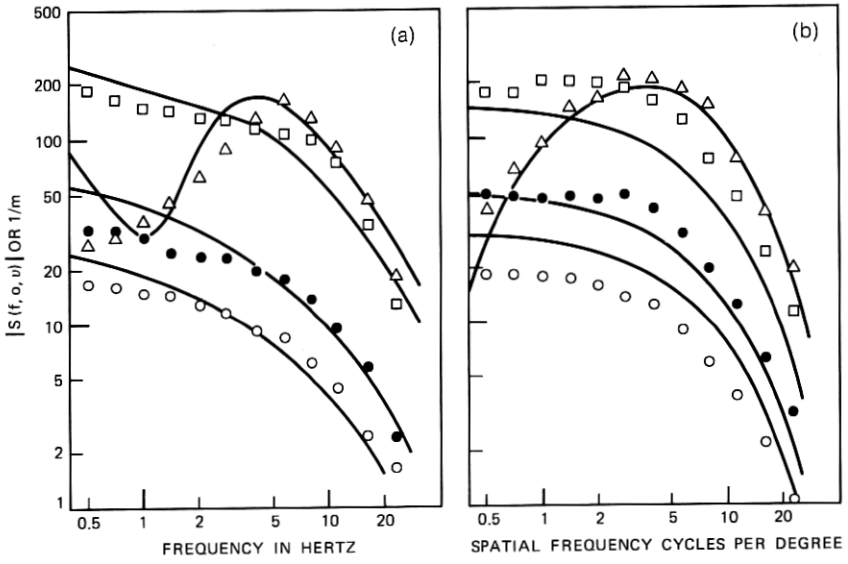


Fig. 9—Diffusion-Gaussian/Cauchy model applied to Robson's data. (a) Temporal frequency response, parameters as in Fig. 5a. (b) Spatial frequency response, parameters as in Fig. 5b.

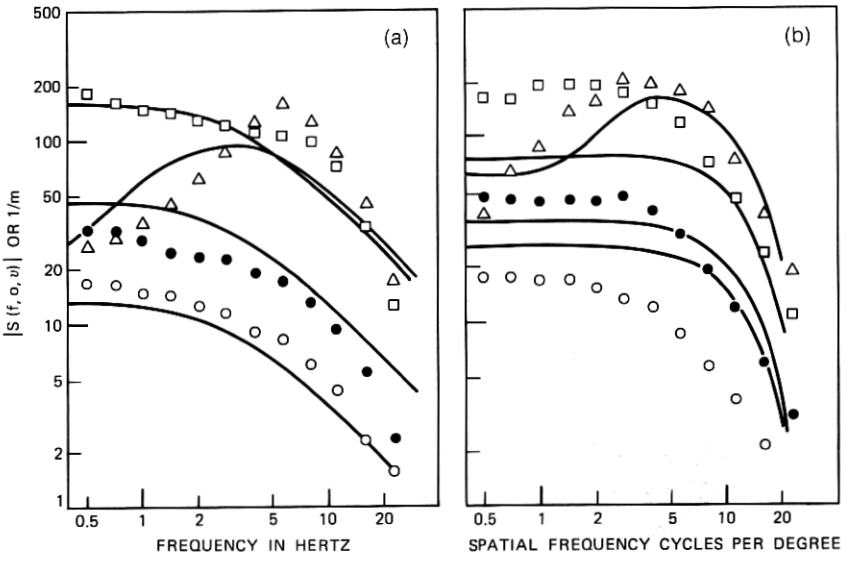


Fig. 10—Poissonian/Gaussian model applied to Robson's data. (a) Temporal frequency response, parameters as in Fig. 5a. (b) Spatial frequency response, parameters as in Fig. 5b.

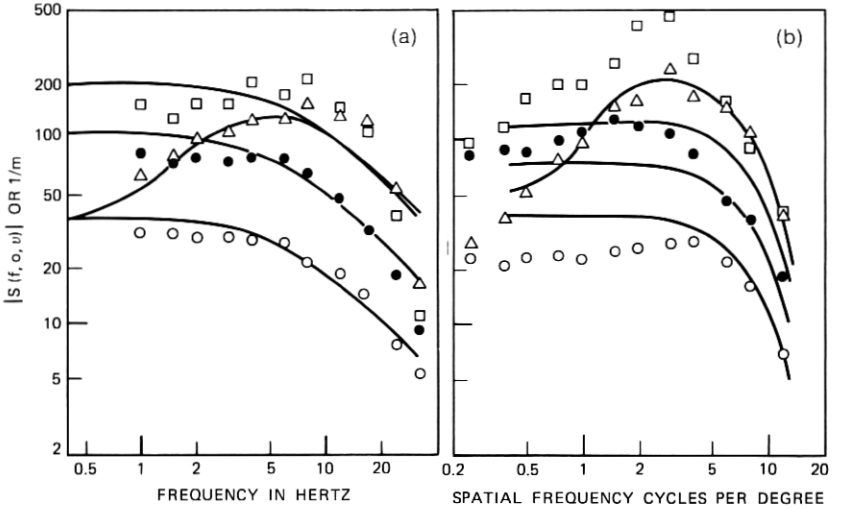


Fig. 11—Poissonian/Gaussian model applied to Kelly's data at 62.8 mL. (a) Temporal frequency response, parameters as in Fig. 1a. (b) Spatial frequency response, parameters as in Fig. 1b.

(k). The time constants tend to increase with lower luminance while the space constants remain unchanged. The parameter values for the D-G/C model obtained using the log of departure criterion are given in Table III. With this different model, parameter values are naturally very different, but the variations with luminance are similar to those with the P/G model and, indeed, with the remaining models.

TABLE IIA—PARAMETER VALUES IN POISSONIAN/GAUSSIAN MODEL DETERMINED WITH ASFE PERFORMANCE INDEX

	Parameter \ Luminance (mL)	Kelly's Data				Robson's Data
		62.8	15.2	3.7	0.91	6.3
1	A	298	236	145	116	219
2	τ_1 (ms)	39	32	32	61	45
3	τ_2 (ms)	63	43	70	102	52
4	σ_e (min arc)	1.48	1.55	1.49	1.49	1.01
5	σ_i (min arc)	9.82	4.72	10.1	6.19	5.62
6	k	0.9976	0.9831	0.9579	0.8150	0.9554
7	D	7.0	4.3	3.9	7.6	3.7
8	A/L_o	4.82	15.5	40.3	127.2	34.8
9	$(1 - k)$	0.0024	0.0169	0.0421	0.1850	0.0446

TABLE IIB—PARAMETER VALUES IN POISSONIAN/GAUSSIAN MODEL DETERMINED WITH ASLE CRITERION [EQ. (35)]

	Parameter \ Luminance (mL)	Kelly's Data				Robson's Data
		62.8	15.2	3.7	0.91	6.3
1	A	234	168	134	93	198
2	τ_1 (ms)	29	37	53	79	55
3	τ_2 (ms)	34	58	80	101	55
4	σ_e (min arc)	1.52	1.40	1.37	1.34	1.01
5	σ_i (min arc)	9.68	8.30	10.51	10.28	5.58
6	k	0.990	0.971	0.911	0.740	0.996
7	D (log units)	0.48	0.56	0.63	0.73	0.49
8	A/L_o	3.72	11.05	36.2	102	31.4
9	$(1 - k)$	0.010	0.029	0.089	0.260	0.004

V. DISCUSSION

Both the D-G/C and the P/G models will be useful in practice, particularly the latter when simplicity of computation is a major consideration. However, the fact that none of the models fits the data well enough to satisfy any fundamental inquiry prompts us to look again at the assumptions of Section 2.1.

One can scarcely doubt the interplay of excitation and inhibition in the visual mechanism, and that inhibition spreads over a wider

TABLE III—PARAMETER VALUES IN DIFFUSION-GAUSSIAN/CAUCHY MODEL DETERMINED WITH ASLE PERFORMANCE INDEX [EQ. (35)]

	Parameter \ Luminance (mL)	Kelly's Data				Robson's Data
		62.8	15.2	3.7	0.91	6.3
1	A	1596	943	810	372	853
2	τ_1 (ms)	472	489	649	656	496
3	τ_2 (ms)	74	75	74	111	98
4	σ_e (min arc)	9.33	8.01	6.47	7.43	8.59
5	σ_i (min arc)	12.38	11.45	6.50	8.27	32.4
6	k	0.517	0.479	0.351	0.236	0.677
7	D (log units)	0.45	0.48	0.50	0.61	0.33
8	A/L_o	25.4	62	219	409	135
9	$(1 - k)$	0.483	0.521	0.649	0.764	0.323

area and persists longer than the excitation, i.e., is confined to lower spatial and temporal frequencies. However, it is probably untrue that inhibition simply subtracts from the excitation. It is more likely[17] that it acts as a shunt, or a reduction in through-put gain, for which simple subtraction is only a first approximation. One could also expect a more precise characterization of inhibitory action to explain part of the adaptive changes. However, the model would be nonlinear and more complicated.

Apart from linearity, it is very probable that more separability of functions has been assumed than is warranted. The statement that excitation (or inhibition) is separable into space and time functions purports that, given a point flash, the form of the spatial response is independent of time, or that the shape of the time function is independent of distance from the stimulus point. This is probably true of the spread which is due to optical smearing of the retinal image. But it is probably untrue of the lateral spread of neural interactions. Since neural interactions predominate in the wider inhibitory spread, separability should be expected to be a poorer assumption for inhibition than for excitation. This seems to be borne out by the data.

The assumption of uniformity raises another question. To speak of isoplanatic patches is, of course, no more than a simplification. Even the central fovea varies substantially in receptor packing density within the space of less than a degree. It is therefore difficult to maintain the assumption of uniformity with data obtained for spatial frequencies of one cycle/degree or lower. To justify convolution in the presence of nonuniformity we only need to be sure that the spatial spread is small compared to the size of the "uniform" patch. However, we need uniformity over much more than $(1/f_c)$ in order to justify a Fourier transform to within f_c of the frequency origin. If this condition is not met, then with a sinusoidal input the output may, in the extreme, be nonsinusoidal even over only a part of a cycle. But our assumption of threshold is that a criterion value be exceeded by the peak-to-peak output and this then will not be related to the calculated transfer function.

The concept of detection needs to be examined, not only where lack of retinal uniformity is critical. It is unlikely that detection is based on a comparison of just two values, a maximum and a minimum in the output, and that this comparison is independent of how far apart in space and time these two values actually are. It is more likely that there should be a pooling of evidence and that there should be a decline in detectability, the further apart the relevant events.

However, it need not follow that, given a more complicated detection mechanism, the modeling done here would be invalidated. The detector with variable weighting of evidence could, in fact, be equivalent to a spatial/temporal filter in its own right, followed by the kind of decision stage assumed here. If this were so, then it would only mean that not all the filtering evident from threshold data can be attributed to peripheral processes, but that some of it is due to central neural activity. This is an important distinction where comparisons are made between the filtering evident in stimulus detection and in, say, perception of brightness. Inconsistencies of this nature have already been noted in the literature,[18] but have not been satisfactorily explained.

Higher-level filtering might also be responsible for the frequency-selective fatiguing discovered by Blakemore and Campbell.[19] It seems improbable that spatial filtering by optical and retinal spread constitutes spatial frequency channels which may be independently adapted, but higher-level filtering could, in fact, occur after a Fourier-like signal transformation. But again, the presence of any transformations like these would not affect the present modeling. They might however, affect adaptation effects.

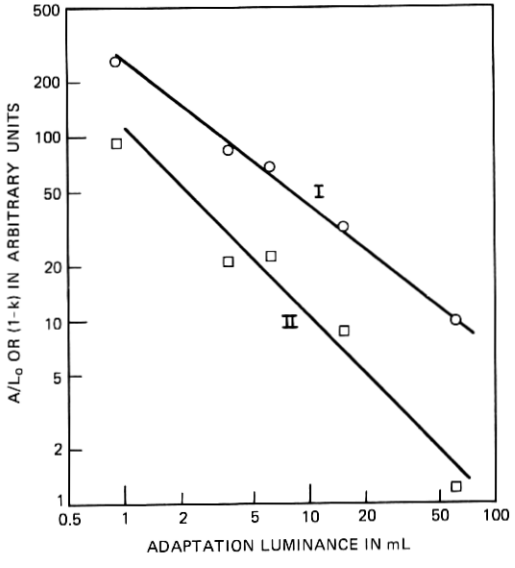


Fig. 12—Adaptation of gain parameters A/L_o (line I) and $(1 - k)$ (line II) against luminance as obtained in fitting Poissonian/Gaussian model to Kelly's and Robson's data.

Of the adaptive changes which are evident from the present modeling, the variation in gain requires comment. The fact that the gain constant A was seen to decrease with decreasing adaptive luminance might be taken to mean that the system becomes less sensitive in the dark. As is well known, visual sensitivity goes up markedly with darkness and the present results do not, in fact, contradict this. By eq. (8), $(1/m)$ equals $|S(u, 0, f)|$ only to within the multiplicative constant $T(p)/L_o$. Assuming that the threshold does not change, then to make the gain values at different luminances L_o comparable to each other they have to be divided by L_o. A/L_o does in fact go up with decreasing luminance as can be seen in row 8 of Table IIA and elsewhere. The actual A/L_o values are different across the models but the trend is always the same.

As the adaptation luminance decreases there is an additional increase in sensitivity restricted to low frequencies. This occurs because of the decline in fractional inhibition k. The zero-point value of $|S|$ is with all models $A(1 - k)/L_o$. The net excitation $(1 - k)$ is given in row 9 of Tables IIA, IIB, and III. A/L_o and $(1 - k)$ have also been plotted for the P/G model in Fig. 12. From the plot one can infer that for the P/G model and in the range $1.0 \leqq L_o \leqq 100$ mL

$$A/L_o = \text{const}_1 \times L_o^{-0.81}, \tag{43}$$

$$(1 - k) = \text{const}_2 \times L_o^{-1.03}, \tag{44}$$

so that

$$|S(0, 0, 0)| = \text{const}_3 \times L_o^{-1.84}. \tag{45}$$

The increase in low-frequency sensitivity with decreasing luminance is at the expense of bandwidth.

VI. CONCLUSION

Six spatio-temporal models of human visual filtering were tested against published experimental data on visual spatio-temporal sine-wave thresholds. These models arose as specific examples from a definite theoretical framework. It was assumed that thresholds could be related to a fixed peak-to-peak difference in a visually filtered version of the input stimulus, and that the filtering could be taken as time-invariant and spatially uniform and isotropic. Particular attention was directed to the question of whether the response was separable into functions of time and space. We showed that the total response is not so separable in this way. However, it was assumed that if the

response is expressed as an algebraic difference of two terms, excitation and inhibition, the individual terms would be separable.

Component functions which were tried were exponential, Gaussian, and diffusion-like functions of time, and Gaussian and Cauchy functions of space. The best fit was obtained with a model which has a diffusion-like time function for excitation, a Gaussian time function for inhibition, and Cauchy space functions for both. The diffusion function, as a model of the time course of excitation, has previously been advocated by Ives,[14] Kelly,[12] and others. The degree of fit obtained in the present study, involving both time and space, was however only moderate and no strong argument can be brought forward in favor of any of the functions, not even the best-fitting. In the best case the average departure from the model was three times larger than the average estimated experimental error. The present results do not exclude any of the functions either, for the fit was probably affected more by the restrictions of the framework than the choice of function.

In each of the models six parameter values had to be determined. These were gain, fractional inhibition, two time constants, and two space constants. Parameter searches consisted of up to 50 passes of gradient-dependent convergence and evolutionary random search. Random search was invariably found to be the more productive phase in all the computational passes.

With adaptation luminance between 1 and 60 mL, the time constants were found to be slightly larger at the low luminances than at the high, the space constants were almost nonvarying, and the gain and fractional inhibition decreased with decreasing luminance. As expected, the sensitivity, measured as gain divided by luminance, was found to go up with decreasing luminance. The reduction in fractional inhibition was shown to give a further increase in sensitivity with decreasing luminance, but only at low frequencies. With one model (P/G) the sensitivity at zero frequency was found to vary inversely as the 1.84 power of luminance, 0.81 of this being due to variation in overall sensitivity and the remainder due to changes in inhibition.

The major purpose of the present model fitting was to find a filter function for use in a program for predicting the subjective quality of visual signal coding schemes. Of the six models the most economical computational procedures are provided by the Poissonian/Gaussian model. The Poissonian, or negative exponential, time functions can be implemented recursively, using a delay of only one or two picture frames, and the Gaussian space functions, being themselves separable

into products of functions of x and y, can be implemented by two successive, modest transverse filter operations, instead of requiring one very large operation. This model was found to fit the data nearly as well as the best. Considering its computational advantages, it will no doubt be the one to find most use.

VII. ACKNOWLEDGMENTS

Both J. J. Robson of the University of Cambridge and D. H. Kelly of the Stanford Research Institute kindly explained their experiments and discussed their data. As noted, Dr. Kelly was also able to send copies of data and estimates of experimental errors.

REFERENCES

1. Ives, H. E., "Critical Frequency Relations in Vision," J. Opt. Soc. Am. and Rev. Sci. Instr., *6*, 1922, pp. 254–268.
2. Schade, O. H., "Optical and Photoelectric Analog of the Eye," J. Opt. Soc. Am., *46*, 1956, pp. 721–739.
3. Kelly, D. H., "J_0 Stimulus Pattern for Visual Research," J. Opt. Soc. Am., *50*, 1960, p. 1115.
4. Robson, J. J., "Spatial and Temporal Contrast-Sensitivity Functions of the Visual System," J. Opt. Soc. Am., *56*, 1966, pp. 1141–1142.
5. Kelly, D. H., "Frequency Doubling in Visual Responses," J. Opt. Soc. Am., *56*, 1966, pp. 1628–1633.
6. Kelly, D. H., "Adaptation Effects on Spatio-Temporal Sine-Wave Thresholds," Vision Res., *12*, 1972, pp. 89–101.
7. Budrikis, Z. L., "Visual Fidelity Criterion and Modeling," Proc. IEEE, *60*, 1972, pp. 771–779.
8. LeGrand, Y., *Light, Colour and Vision*, London: Chapman and Hall, 1968 (2nd Edition), p. 106.
9. Linfoot, E. H., *Fourier Methods in Optical Image Evaluation*, London and New York: The Focal Press, 1964, p. 15.
10. Westheimer, G., and Campbell, F. W., "Light Distribution in the Image Formed by the Living Human Eye," J. Opt. Soc. Am., *52*, 1962, pp. 1040–1045.
11. Ratliff, F., Hartline, H. K., and Miller, W. H., "Spatial and Temporal Aspects of Retinal Inhibitory Interaction," J. Opt. Soc. Am., *53*, 1963, pp. 110–120.
12. Kelly, D. H., "Theory of Flicker and Transient Responses, I. Uniform Fields," J. Opt. Soc. Am., *61*, 1971, pp. 537–546. Reference is given to earlier papers where high-frequency asymptote was first pointed out.
13. Graham, N. Y., "Perspective Drawing of Surfaces With Hidden Line Elimination." B.S.T.J., *51*, No. 4 (April 1972), pp. 843–861.
14. Ives, H. E., "A Theory of Intermittent Vision," J. Opt. Soc. Am. and Rev. Sci. Instr., *6*, 1922, pp. 343–361.
15. Roberts, G. E., and Kaufman, H., *Table of Laplace Transforms*, Philadelphia and London: Saunders, 1966, p. 246.
16. Cantoni, A., "Optimal Approximations with Piecewise Linear Functions," Ph.D. Thesis, The University of Western Australia, 1972.
17. Sperling, G., "Model of Visual Adaptation and Contrast Detection," Perception Psychophys., *8*, 1970, pp. 143–157.
18. Hay, G. A., and Chester, M. S., "Signal-Transfer Functions in Threshold and Super Threshold Vision," J. Opt. Soc. Am., *62*, 1972, pp. 990–998.
19. Blakemore, C., and Campbell, C. W., "Adaptation to Spatial Stimuli," Proc. Physiol. Soc., September 1968, pp. 11p–13p.