Temporal covariance model of human motion perception

Jan P. H. van Santen and George Sperling

Department of Psychology, New York University, 6 Washington Place, New York, New York 10003

We propose a model of direction-sensitive units in human vision. It is a modified and elaborated version of a model by Reichardt [Z. Naturforsch. Teil B **12**, 447 (1957)]. The model is applied to threshold experiments in which subjects view adjacent vertical bars with independently (typically sinusoidally), temporally modulated luminances. The subject must report whether the patterns moved to the left or to the right. According to the model, a basic motion-detecting unit consists of two subunits tuned to opposite directions. Each performs a spatial and temporal linear filtering of its input; outputs of the filters are multiplied, and the multiplied output is integrated (for a time that is long relative to the modulation period). The model's output consists of the difference between the subunit outputs. Direction of movement is indicated by the sign of the model output. Mathematical analysis of the model yielded several predictions that were confirmed experimentally. Specifically, we found that (1) performance with complex patterns can be predicted by spatiotemporal Fourier analysis that results in the segregation and linear addition in the output for different temporal frequencies; (2) under special conditions, performance depends on the product of adjacent bar amplitudes, offering strong support for the multiplication principle; (3) performance is unaffected by addition of stationary patterns; and (4) addition of homogeneous flicker normally produces no effect but under special conditions reverses perceived direction. These and other results confirm our model and reject several other models, including Reichardt's original model.

INTRODUCTION

Moving objects form a complex, spatiotemporal stimulation pattern. In a large variety of situations, the human visual system manages to infer both direction and amount of movement from this pattern. What are the algorithms by which our visual system accomplishes this task?

There is a good chance that this question does not have a general answer. That is, there is widespread agreement that the human visual system contains at least two motion analysis systems.[1-4] The first system, the short-range process[1] is assumed to operate primarily under conditions of temporally and spatially continuous stimulation. The second system operates in classical apparent-motion situations, in which wide spatial and temporal intervals separate successive stimulations. The typical assumption is that the first system consists of simple mechanisms early in the visual system and that the second system involves higher-level processes. We are concerned here with the first system.

Campbell and Robson,[5] Wilson and Bergen,[6] and many others have proposed that the human visual system contains channels that have spatiotemporal frequency selectivity. Moreover, these channels appear to perform a Fourier analysis in the sense that one can predict psychophysical responses quite well on the basis of a three-dimensional (two spatial and one temporal dimension) Fourier analysis of displays and by considering the dominant component.[7-9] So far, the work on frequency-selective channels has not led to the development of models describing mechanisms that perform the Fourier analysis of moving stimuli. In this paper, we propose such a model. It is a modified and elaborated version of a model developed in the context of experiments on insects by Reichardt and his collaborators.[10]

Outline

We discuss the following in order: (1) our stimuli; (2) Reichardt's original model; (3) our modifications and elaborations

of it; (4) several alternative models (developed mostly in artificial intelligence research); (5) Experiments 1–3, which serve primarily as demonstrations of the original and the elaborated Reichardt models; (6) Experiment 4, which shows the necessity of our modifications and elaborations; and (7) Experiments 5–7, which test the original and the elaborated Reichardt model against the other models.

1. MOTION ANALYSIS MODELS

A. Basic Motion Display

We give here a brief description of our displays, which are patterns of white light produced on a computer controlled cathode-ray tube (CRT) that is viewed binocularly in a dimly lighted room. Let the time-varying two-dimensional display be $L(x, y, t)$, where L is the luminance (gray level) at a location with spatial coordinates x and y [in degrees of visual angle (in degrees)] at time t (in seconds). Our displays have the following properties.

First, they are unidimensional and move in the horizontal direction. That is, $L(x, y, t)$ is a function only of x and t within the viewing window and zero otherwise. For short, we will write $L(x, t)$.

Second, $L(x, t)$ is periodic with time, with period T sec. An important implication is that for fixed x, $L(x, t)$ has a Fourier series decomposition in the time dimension, $\sum_{n=0}^{\infty} \zeta_n \sin(2\pi n\omega t - \eta_n)$. Here, $\omega = 1/T$, and ζ_n and η_n are the usual Fourier series coefficients. Taking into account that, in general, ζ_n and η_n depend on the spatial coordinate x, we can write $L(x, t) = \sum_{n=0}^{\infty} L_n(x, t)$, where

$$L_n(x, t) = \zeta_n(x)\sin[2\pi n\omega t - \eta_n(x)]. \tag{1}$$

We will refer to each $L_n(x, t)$ as a temporal-frequency component of $L(x, t)$ with temporal frequency $n\omega$ Hz.

Third, displays in our experiments consist of identically shaped parallel vertical bars, $B_1, \ldots, B_n$, each having its own
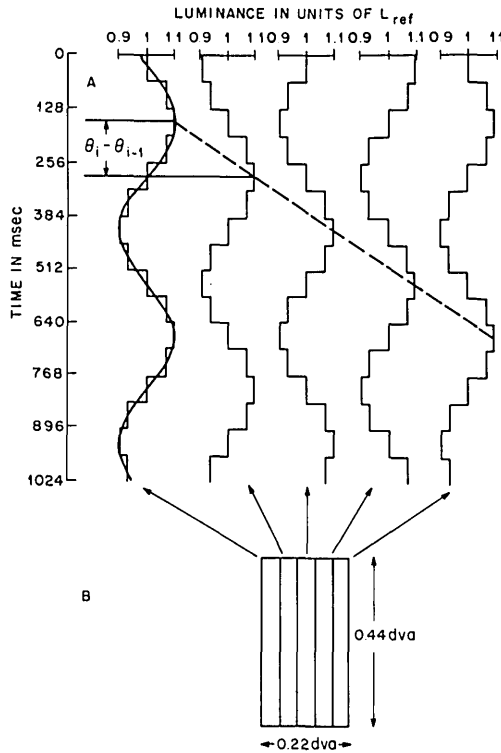
Fig. 1. Representation of a five-field display. A, luminance modulations in each of five fields, $L_j(t) = L_{0j} + m_j h(t - \vartheta_j)$, is here represented as $L_j(t)/L_{ref}$, where $L_{ref} = 51$ cd/m^2 for most displays. The modulation function $h(t)$ is an 8-point approximation to a sinusoid, which is indicated by a continuous curve for the leftmost field. Between-field asynchrony in this example is 128 msec (1/4 cycle, $\pi/2$ rad). The temporal phase line (dashed line) interconnects the peak luminances of the dominant Fourier component in each bar. B, the five-field display. Overall height and width are 0.44 and 0.22 deg, respectively.

luminance modulation function $L_j(t)$. More formally, let $[b_{j-1}, b_j]$ be the x interval occupied by bar B_j. Then $L(x, t) = L_j(t)$ if $x \in [b_{j-1}, b_j]$. In addition, the functions $L_j(t)$ are always of the form

$$L_j(t) = L_{0j} + m_j h(t - \vartheta_j). \qquad (2)$$

In other words, the same time-periodic modulation function $h(t)$ occurs in each bar, but with different mean luminance L_{0j}, amplitude m_j, and phase ϑ_j.

Figure 1 represents a display in which $L_{0j} = 100$ luminance units, $m_j = 10$ units, and $h(t)$ is the sine function with a period T of 512 msec (hence temporal frequency ω of 1.95 Hz) and a between-bar asynchrony $\vartheta_{j+1} - \vartheta_j$ of 128 msec ($T/4$). When the between-bar asynchrony is the same for all pairs of adjacent bars, it will be called the phase difference of a display, denoted φ in radians; in the figure, $\varphi = \pi/2$.

The time periodicity of $h(t)$ implies that it has the Fourier series expansion

$$h(t) = \sum_{n=0}^{\infty} \varepsilon_n \sin(2\pi n \omega t - \kappa_n). \qquad (3)$$

When h is dominated by only one Fourier component (with temporal frequency ωn_0), there is a simple heuristic for inferring direction of motion from stimulus representations of the type depicted in Fig. 1. Consider the dashed line in Fig. 1 that interconnects the nearest peak luminances of the

dominant Fourier component in each bar (indicated by a continuous curve for the first bar). We call this line the temporal phase line. The sign and (inverse) magnitude of its slope are indices of direction and velocity of motion: rightward motion leads to a temporal phase line that descends (from left to right), and fast motion leads to a shallow line. Below, we shall see that the directional response of a simplified version of the Reichardt model can be predicted on the basis of the temporal phase line.

B. Reichardt's Model

The original Reichardt model[10] contains temporal filters that are irrelevant to the predictions tested in this paper. Therefore we consider a slightly simplified version of Reichardt's model.

According to Reichardt, a motion detector consists of two subunits that are mirror images of each other (Fig. 2A). The input to these subunits consists of $L(x, t)$ sampled at locations x_{left} and x_{right}, i.e., $L(x_{left}, t)$ and $L(x_{right}, t)$. In other words, the model assumes that input channels have point receptive fields that are shared by the two subunits.

The subunits are tuned to motion in opposite directions. Output from the right subunit, which is tuned to rightward motion, reflects how well $L(x_{left}, t)$, after passing through a linear temporal filter, matches $L(x_{right}, t)$. The left subunit does the corresponding operation.

The final output of the detector is given by the difference between the subunit outputs. The sign of this difference indicates direction of motion. For now we leave it open whether this subtraction operation reflects a conscious process of comparing the subunit outputs or a preconscious inhibitory process in which only the final result is accessible to consciousness.

Before we give a detailed description of the original Reichardt model, we would like to point out three potential obstacles that the reader might have in developing an intuition for it. The first is that one usually thinks of motion as involving a spatial object that occupies different locations at different points in time. In the model, one has to do the reverse—think of motion as involving a temporal object (luminance modulation pattern) that occurs at different points in time in different locations. The second is that a linear temporal filter does more than perform a simple delay operation. It delays different temporal modulation functions by different amounts and, in general, alters their shapes. The third is that the output of the detector does not simply reflect how well $L(x_{left}, t)$ after filtering matches $L(x_{right}, t)$, or vice versa. Rather, because of the subtraction operation, detector output reflects the between-subunit difference in well-matchedness.
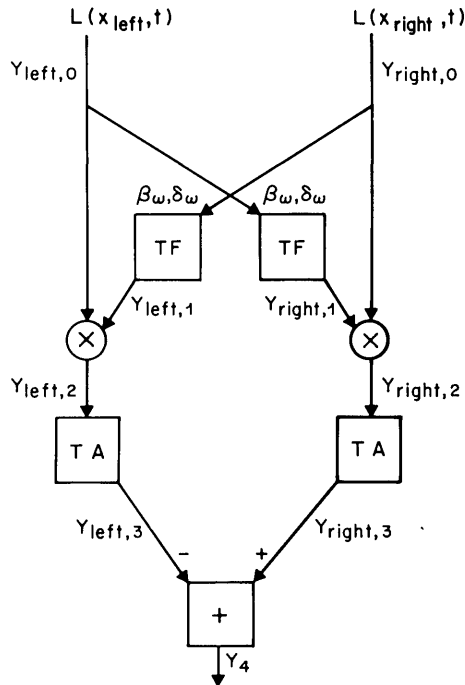
We now give a detailed description of the original Reichardt model. We denote the signal at various levels of processing in the unit with $y_{H,j}$. The H (think of H and) takes the values left and right, whereas j refers to the level of processing.

The signal in the left (right) input channel $y_{H,0}(t)$ is simply $L(x, t)$ sampled at the point x_H. This can be written as $\sum_{n=0}^{\infty} L_n(x_H, t)$, where $L_n(x, t)$ is defined in Eq. (1). It is easy to show that $L_n(x_H, t)$ is of the form $\alpha_{H,n} \sin(2\pi n \omega t - \gamma_{H,n})$. Hence

$$y_{H,0}(t) = L(x_H, t) = \sum_{n=0}^{\infty} \alpha_{H,n} \sin(2\pi n \omega t - \gamma_{H,n}).$$

If we adopt the convention that $\gamma_{H,0} = \pi/2$, $\alpha_{H,0}$ is the mean
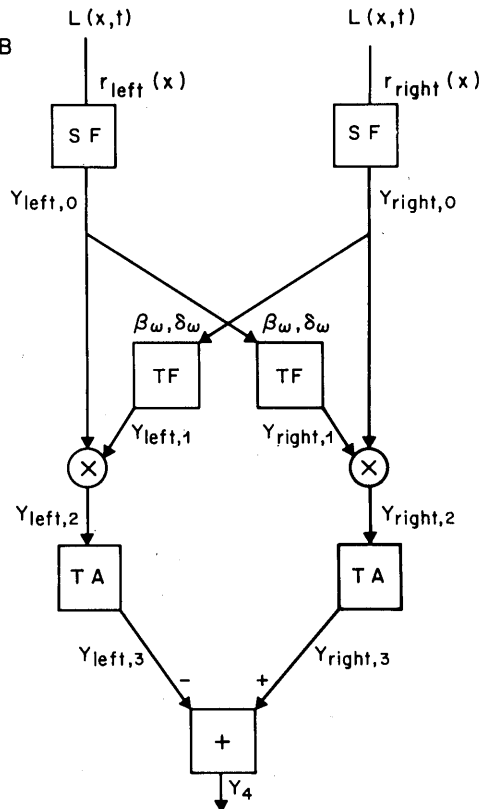
A

$L(x_{left}, t)$ $L(x_{right}, t)$

$Y_{left,0}$ $Y_{right,0}$

$\beta_\omega, \delta_\omega$ $\beta_\omega, \delta_\omega$

TF TF

$Y_{left,1}$ $Y_{right,1}$

$\times$ $\times$

$Y_{left,2}$ $Y_{right,2}$

T A T A

$Y_{left,3}$ $Y_{right,3}$

$-$ $+$

$+$

Y_4

B

$L(x,t)$ $L(x,t)$

$r_{left}(x)$ $r_{right}(x)$

S F S F

$Y_{left,0}$ $Y_{right,0}$

$\beta_\omega, \delta_\omega$ $\beta_\omega, \delta_\omega$

TF TF

$Y_{left,1}$ $Y_{right,1}$

$\times$ $\times$

$Y_{left,2}$ $Y_{right,2}$

T A T A

$Y_{left,3}$ $Y_{right,3}$

$-$ $+$

$+$

Y_4

Fig. 2. A, the Reichardt model. Input consists of the stimulus $L(x, t)$ sampled at locations x_{left} and x_{right}; $y_{i,H}$ represents the signal at the various stages i for the left and right subunits (H = left, right). TF indicates a linear, time-invariant filter with attenuation β_ω and phase shift δ_ω; $\times$ indicates a multiplication unit; TA indicates a time averaging unit, and $+$ indicates a unit that adds its (negative and positive) inputs. B, proposed modification of the Reichardt model in which the point input assumption is generalized to the input of the entire stimulus $L(x, t)$ through a linear spatial filter, denoted SF.

luminance of $L(x_H, t)$ and

$$y_{left,0}(t) = \alpha_{H,0} + \sum_{n=1}^{\infty} \alpha_{H,n} \sin(2\pi n\omega t - \gamma_{H,n}). \qquad (4)$$

Two operations are performed on the input. The first is linear time-invariant filtering (marked TF in Fig. 2A). This type of filtering has the following well-known properties. First, the response of a linear filter to the sum of two inputs is equal to the sum of the responses to the separate inputs. Second, the response of a linear filter to $\alpha_{H,n} \sin(2\pi n\omega t - \gamma_{H,n})$ is $\alpha_{H,n}\beta_{n\omega} \sin(2\pi n\omega t - \gamma_{H,n} - \delta_{n\omega})$. In other words, a linear time-invariant filter affects only amplitude and phase of a sine wave and does not alter the waveform. Sensitivity $\beta_{n\omega}$ and phase shift $\delta_{n\omega}$ of a linear filter depend only on temporal frequency ($n\omega$). When we let $n = 0$, we see that a constant input $\alpha_{H,0}$ corresponds to a constant output time function $\beta_0\alpha_{H,0}$. It follows from these considerations that the output $y_{H,1}$ of the linear filter is $\beta_0\alpha_{H',0} + g_{H'}(t)$, where

$$g_H(t) = \sum_{n=1}^{\infty} \alpha_{H,n}\beta_{n\omega} \sin(2\pi n\omega t - \gamma_{H,n} - \delta_{n\omega}). \qquad (5)$$

The other operation on the input (marked $\times$ in Fig. 2A) is multiplication of $y_{left,0}(t)$ with $y_{left,1}(t)$ in the left subunit. After application of some basic trigonometric rules, we see that the result is

$$y_{left,2}(t) = \beta_0\alpha_{left,0}\alpha_{right,0}$$

$$+ \alpha_{left,0} \sum_{m=1}^{\infty} \alpha_{right,m}\beta_{m\omega} \sin(2\pi m\omega t - \gamma_{right,m} - \delta_{m\omega})$$

$$+ \alpha_{right,0}\beta_0 \sum_{m=1}^{\infty} \alpha_{left,n} \sin(2\pi n\omega t - \gamma_{left,n})$$

$$+ \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{2} \alpha_{left,n} \alpha_{right,m}\beta_{m\omega}$$

$$\times \{\cos[2\pi(n - m)\omega t - \gamma_{left,m} + \gamma_{right,m} + \delta_{m\omega}]$$

$$- \cos[2\pi(n + m)\omega t - \gamma_{left,m} - \gamma_{right,m} - \delta_{m\omega}]\}. \qquad (6)$$

For the right subunit we interchange left and right everywhere in Eq. (6).

The next operation is time averaging (marked TA in Fig. 2A). Here, we define the time average $y_{H,3}(t)$ of the function $y_{H,2}$ as $\lim_{K \to \infty} 1/(t + K) \int_{-K}^{t} y_{H,2}(\tau) \, d\tau$. Note that we have defined the time average as the average from the (infinite) past up to the present. In principle, the time average is a function of t (the present time); however, for periodic input, the time average does not vary. Time averaging has three properties. First, the time average of a sine or cosine is zero. Second, the time average of a constant function equals its constant value. Third, time averaging is a linear operator. Time averaging eliminates all time-dependent components in Eq. (6) because these components ultimately are expressible as sums of simple sines and cosines, each one of which vanishes. The only components that are not time dependent are the first term and some terms in the double sum that arise when $n = m$. Hence Eq. (6) reduces to

$$y_{left,3}(t) = \beta_0\alpha_{left,0}\alpha_{right,0} + \sum_{n=1}^{\infty} \frac{1}{2} \alpha_{left,n} \alpha_{right,n}\beta_{n\omega}$$

$$\times \cos[\delta_{n\omega} - (\gamma_{left,n} - \gamma_{right,n})]. \qquad (7)$$

Again, for the right subunit, left and right must be interchanged. Note that $y_{left,3}$ is time independent. Equation (7)

is the output from the left subunit. Its value is maximal when $\gamma_{\text{left},\omega} - \gamma_{\text{right},\omega}$ (the right-left temporal phase difference) matches δ_ω (the subunit's temporal phase delay).

The final subtraction operation results in cancellation of the term $\beta_0 \alpha_{\text{right},0} \alpha_{\text{left},0}$, which arises from the mean luminance. Using elementary trigonometry, we can now write the (time-independent) response to input $L(x, t)$ as

$$y_4(L) = \sum_{n=1}^{\infty} \alpha_{\text{right},n} \alpha_{\text{left},n} \beta_{n\omega} \sin\delta_{n\omega} \sin(\gamma_{\text{right},n} - \gamma_{\text{left},n}).$$

(8a)

The response to $L_n(x, t)$ is

$$y_4(L_n) = \alpha_{\text{right},n} \alpha_{\text{left},n} \beta_{n\omega} \sin\delta_{n\omega} \sin(\gamma_{\text{right},n} - \gamma_{\text{left},n}).$$

(8b)

Note that we include L (or L_n) as an argument of y_4. Strictly speaking, we should have done the same for the α_n's and γ_n's, since they also depend on L (i.e., on L_n).

Because of its multiplication component, the model is highly nonlinear. Remarkably, Eqs. (8) imply that the response of a detector to $L(x, t)$ is the sum over n of the responses to the temporal-frequency components $L_n(x, t)$. We designate this as the property of segregation of temporal frequencies. This property can be considered a weak version of linearity; it is weak because it asserts linearity only for temporal-frequency components of L, whereas true linearity would require the sum rule to apply to any decomposition of L (including, for example, spatial-frequency components). A further implication of segregation of temporal frequencies is that detector output is not affected by changes in the time-independent component of $L(x, t)$, $L_0(x, t)$. This follows from the fact that Eq. (8a) does not contain $\alpha_{\text{left},0}$ or $\alpha_{\text{right},0}$, thus implying that the response to $L_0(x, t)$ is zero.

Our equation for detector output [Eq. (8a)] is different in two respects from Reichardt's. First, we deal with any time-periodic function $L(x, t)$, whereas Reichardt deals only with rigid motion under uniform illumination. In the latter case, displays can be represented by functions of the type $L(x, t)$ that have the property $L(x_{\text{right}}, t) = L(x_{\text{left}}, t - \Delta t)$, where Δt is the time it takes for the object to move from x_{left} to x_{right}. As we will demonstrate in this paper, applicability to nonrigid motion is critical for rigorous empirical tests of the model. Second, although Reichardt reported results from an empirical test of an implication of the property of segregation of temporal frequencies, namely, relative phase invariance, Reichardt's formulation did not highlight this property itself. Relative phase invariance refers to the property that for a shift of each L_n backward or forward in time by arbitrary amounts, $y_4(L)$ does not change. This property follows directly from the segregation property, together with the obvious fact that, for any n, $y_4(L_n)$ is unaffected by changes in the onset of L_n.

C.   Modifications of the Reichardt Model

As it now stands, the original Reichardt model has severe difficulties in accounting for human motion perception, because the model is vulnerable to a form of aliasing that is exhibited little, if at all, by intact humans or isolated neurons. The problem is, briefly, that as one changes the temporal or spatial frequency, but not the direction, of a moving sine wave,

the sign of the detector response (y_4) reverses. Here, we define a moving sine wave as

$$S(m, d, f, \omega) = L_0 + m\sin(2\pi dfx + 2\pi \omega t + \Phi),$$

(9)

where m is the modulation depth, d ($= -1, 0, 1$) is the direction of movement, f is the spatial frequency, ω is the temporal frequency, and Φ is the initial phase at $x = 0, t = 0$. By substituting S for L in Eq. (8a), it can easily be shown that $y_4[S(m, d, f, \omega)] = m^2 d\beta_\omega \sin\delta_\omega \sin[2\pi f(x_{\text{right}} - x_{\text{left}})]$. Susceptibility to aliasing follows from that fact that nothing in the original Reichardt model prevents the terms $\sin(\delta_\omega)$ and $\sin[2\pi f(x_{\text{right},1} - x_{\text{left},1})]$ from being negative. When the former term is negative, we speak of temporal aliasing, and when the latter term is negative, of spatial aliasing. Either form of aliasing means that detector output can be completely wrong: negative for sine waves that move to the right and positive for sine waves that move to the left. It will turn out, however, that aliasing problems can be prevented by adding a few simple assumptions to the original Reichardt model.

1.   Temporal Aliasing

What we need to prevent temporal aliasing is $\beta_\omega \sin(\delta_\omega) \geq 0$. This can be accomplished either if the phase delay δ_ω is always between 0 and π or if, at those temporal frequencies at which the phase shift δ_ω exceeds π, the sensitivity β_ω is zero. The first option is implemented by assuming, as Reichardt suggested in a special case of his model, that TF is a first-order low-pass filter with weighting function $e^{-t/\tau}$. Absence of temporal aliasing is guaranteed for all ω, because in this case $\delta_\omega = \tan^{-1}(\omega\tau)$, which is always between 0 and $\pi/2$.

2.   Spatial Aliasing

The original Reichardt model gives the wrong response when $\sin[2\pi f(x_{\text{right},1} - x_{\text{left},1})] < 0$. This situation occurs when, for example, the spatial period is somewhat smaller than twice the distance between the input channels, because then $2\pi f(x_{\text{right}} - x_{\text{left}}) > \pi$. Since the point input channels have flat spatial-frequency spectra, we cannot deal with spatial aliasing in the same manner as we dealt with temporal aliasing, namely, by assuming that sensitivity to sine waves, the spatial periods of which are less than $x_{\text{right}} - x_{\text{left}}$, is zero. We propose to make an obvious but important generalization of the original Reichardt model, namely, addition of linear spatial input filters (SF's in Fig. 2B). In the face of what is known about the spatial properties of direction-sensitive mechanisms in mammalian vision, the notion that input channels have extended receptive fields rather than point inputs is, of course, quite reasonable. We now spell out the details of how extended receptive fields can prevent spatial aliasing.

Spatially linear input channels. Our first assumption about the spatial properties of the input channels is that each input is given by

$$y_{H,0} = \int r_H(x)L(x, t)\mathrm{d}x, \qquad H = \text{left, right}.$$

(10)

Here, r_H is the receptive field[11] for input channel H. Note that the original Reichardt model uses a special case of Eq. (10), where $r_H(x)$ is the delta function $\delta(x - x_H)$.

Preventing spatial aliasing. Under the assumption of receptive-field linearity [Eq. (10)], the response of a detector with spatially linear input channels to a moving sine wave is

an immediate generalization of the response of the original Reichardt model. It can be shown that the response is given by

$$y_4[S(m, d, f, \omega)] = m^2 d\beta_\omega \sin\delta_\omega P_{\text{left}}(f)P_{\text{right}}(f)D(f). \quad (11)$$

Here, $P_H(f)$ is the spatial frequency response of $r_H(x)$ defined as $|\int_{-\infty}^{\infty} r_H(x)e^{ifx} \, dx|$; and $D(f)$ is a spatial-frequency dependent factor that depends on the spatial relation between the two receptive fields. For the original Reichardt model, P_{left} and P_{right} have constant values that do not depend on f, whereas $D(f) = \sin[2\pi f(x_{\text{right}} - x_{\text{left}})]$. What receptive-field pairs have the property that the factor $P_{\text{left}}(f)P_{\text{right}}(f)D(f)$ is nonnegative for all f?

Symmetric/antisymmetric receptive-field pairs. Figure 3A shows a left and right receptive field that has the general shape

$$r_{\text{left}}(x) = W(x - x_c)\cos[f_0(x - x_c)],$$

$$r_{\text{right}}(x) = W(x - x_c)\sin[f_0(x - x_c)]. \quad (12)$$

Here, x_c represents the location of the detector and f_0 determines the spatial period of the receptive field. $W(x)$ is symmetric around zero, nonnegative, and decreases as x departs from zero. In Fig. 3A, $W(x)$ is $\exp(-x^2/4\sigma^2)$, where $\sigma = 1.41$, $x_c = 0$, and $f_0 = 1$. These receptive fields are not arbitrarily chosen. First, Gabor[12] showed that these functions, like sine waves in Fourier analysis, can be used to describe any function. Second, Marcelja[13] noted that receptive-field profiles of cells in the mammalian visual cortex fit these functions quite well. This should come as no surprise, since others[14] have shown that symmetric receptive fields can be fitted well by a difference of two Gaussian curves, which for the typical ratio of σ's used can be approximated quite closely by functions given by r_{left} in Eq. (12) (Fig. 4). Third, Pollen and Ronner[15] have found, in cats, pairs of adjacent simple cells, the receptive fields of which are described by Eq. (12). It can be shown that when W has a nonincreasing power spectrum (Gaussian spectra have this property), the detector response is given by

$$m^2 d\beta_\omega \sin \delta_\omega P_{\text{left}}(f)P_{\text{right}}(f). \quad (13)$$

Note that $D(f) \equiv 1$. Since, by definition, $P_H(f) \geq 0$, the arrangement given by Eq. (12) prevents aliasing for all spatial frequencies. From an optimality point of view, this constitutes a strong argument in favor of this receptive-field arrangement.
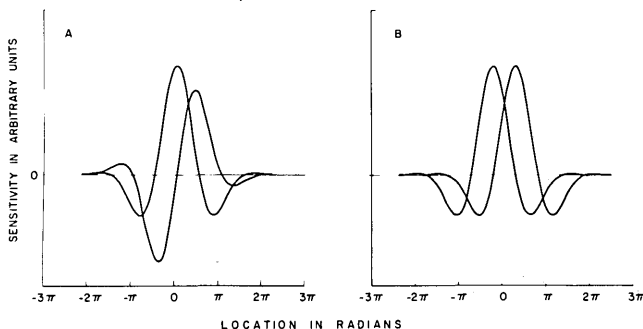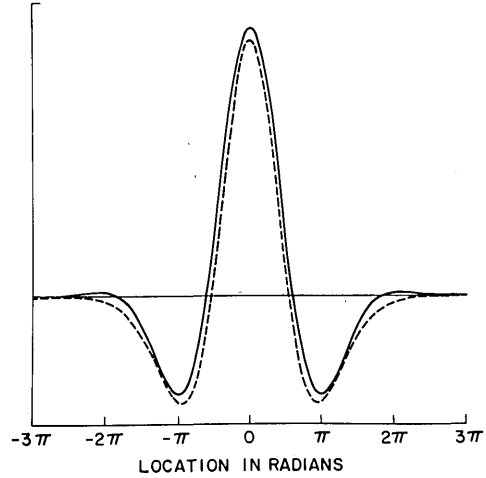

Fig. 4. Comparison of difference of Gaussian curves (dashed curve) and best-fitting (in terms of maximal squared deviation) curves of the type $W(x)\cos(fx)$ (solid curve). Here, $W(x) = A\exp(-x^2/\sigma^2)$, where $\sigma = 2.83$, $A = 1.05$, and $f = 1.76$.

Displaced, symmetric receptive-field pairs. Figure 3B contains a pair of receptive fields that have the general shape

$$r_H(x) = W(x - x_H)\cos[f_0(x - x_H)], \quad H = \text{left, right.} \quad (14)$$

Here, W and f_0 are as in Eq. (12). In Fig. 3B, $x_{\text{left}} = -\pi/2$, $x_{\text{right}} = \pi/2$, $W(x)$ is as in Fig. 3A, and $f_0 = 1$. Because of the symmetry of this receptive-field arrangement it could be considered as being the most direct extension of the original model. Detector response is given by[16]

$$md\beta_\omega \sin\delta_\omega P_{\text{left}}(f)P_{\text{right}}(f)\sin[2\pi f(x_{\text{right}} - x_{\text{left}})]. \quad (15)$$

Here, $D(f) = \sin[2\pi f(x_{\text{right}} - x_{\text{left}})]$. Hence the arrangement in Eq. (14) prevents aliasing only if $P_{\text{left}}(f)P_{\text{right}}(f) = 0$ for all $f \geq [2(x_{\text{right}} - x_{\text{left}})]^{-1}$. This property is only approximately satisfied when W is Gaussian, because its spatial-frequency response is never quite zero.

Implications of spatial linearity of input channels. Input channel linearity [Eq. (10)] allows us to derive several results that are independent of the precise shape of the receptive fields.

(1) *The property of segregation of temporal frequencies is undisturbed.* The most important result is that the assumption of spatial linearity of input channels does not affect the property of segregation of temporal frequencies. To prove this, it can be shown that $\int r_H(x)L_n(x, t)dx$ yields input of the same general form as $L_n(x_H, t)$, i.e., $\alpha_{H,n} \sin(2\pi n\omega t - \gamma_{H,n})$, where the α's and γ's now depend in complicated ways on the receptive-field arrangement. This also implies that detector output is independent of $L_0(x, t)$.

(2) *Derivation of model output for bar input stimuli.* For bar input stimuli, we need a few new terms. Let a_{Hj} be the area under $r_H(x)$ in the interval $[b_{j-1}, b_j]$ occupied by bar B_j: $a_{Hj} = \int_{b_j}^{b_{j+1}} r_H(x)dx$. Let

$$A_{jk} = a_{\text{left},j}a_{\text{right},j+k} - a_{\text{left},j+k}a_{\text{right},j}. \quad (16)$$

Loosely speaking, A_{jk} indicates how well the left and right receptive fields of a detector are differentially aimed at bars

Fig. 3. Candidates for spatial input filters (SF's) in elaborated Reichardt model. A, symmetric–antisymmetric receptive-field arrangement [Eq. (12)]. B, completely symmetric receptive-field arrangement [Eq. (14)].

B_j and B_{j+k}, respectively. To illustrate, consider the original Reichardt model. Here, since $r_H(x) = \delta(x - x_H)$, $A_{jk} = 1$ if $x_{\text{left}} \in B_j$ and $x_{\text{right}} \in B_{j+k}$, and $A_{jk} = 0$ otherwise (note that since $x_{\text{left}} < x_{\text{right}}$ we never have $x_{\text{left}} \in B_{j+k}$ and $x_{\text{right}} \in B_j$). Hence A_{jk} is 1 for those detectors that look at bars B_j and B_{j+k} and zero for all other detectors. The meaning of A_{jk} becomes somewhat more difficult to grasp when we deal with more-complicated receptive fields.

Let $L(x, t)$ be a time-periodic function of the form described in Eq. (2). By going through steps similar to those that led to Eq. (8a), it can be shown that detector response y_4 is

$$y_4(L) = \sum_{k=1}^{F-1} \sum_{j=1}^{F-k} m_j m_{j+k} p(\vartheta_{j+k} - \vartheta_j) A_{jk}. \qquad (17)$$

Here, $p(\vartheta_{j+k} - \vartheta_j) = \sum_{n=1}^{\infty} \varepsilon_n{}^2 \beta_{n\omega} \sin\delta_{n\omega} \sin[2\pi n\omega(\vartheta_{j+k} - \vartheta_j)]$, where $\beta_{n\omega}$ and $\delta_{n\omega}$ are as in Eq. (5), the modulation depth of the nth harmonic ε_n is as in Eq. (3), and the initial temporal phases ϑ_j and ϑ_{j+k} are as in Eq. (2).

(3) *Displays with* $\varphi = \pi/2$. Except for the display used in Experiment 1, our displays fall into two classes. In the first class, $\vartheta_{j+k} - \vartheta_j = \text{const.} = T/4$, so that $\varphi = \pi/2$. It follows that $L_j(t)$ and L_{j+4i} are identical for $i = \pm 0, \pm 1, \pm 2, \dots$ and $j = 1, 2, 3, \dots$. In addition, it follows that L_j and L_{j+4i+2} have the property that corresponding Fourier components have phase differences that are multiples of π: These pairs of bars are in counterphase. Hence bars that are an even number of bars apart do not contain directional information because they are either in phase or in counterphase. The elaborated Reichardt model behaves in accordance with this analysis, because it can be shown that Eq. (17) reduces to

$$y_4(\varphi = \pi/2) = \sum_{k=1,3,5,\dots}^{F-1} \sum_{j=1}^{F-k} m_j m_{j+k} p(\vartheta_{j+k} - \vartheta_j) A_{jk}. \qquad (18)$$

Equation (18) is of particular importance for displays in which the odd-numbered bars have amplitude m_{odd} and the even-numbered bars have amplitude m_{even}. Because k in Eq. (18) is always odd, $j + k$ is even when j is odd, and vice versa; it thus follows that the equation can be written as

$$y_4(\varphi = \pi/2) = m_{\text{odd}} m_{\text{even}} \sum_{k=1,3,5,\dots}^{F-1} \sum_{j=1}^{F-k} p(\vartheta_{j+k} - \vartheta_j) A_{jk}. \qquad (19)$$

This is a powerful equation, since it asserts that, for any display composed of even bars with amplitude m_{even} and odd bars with amplitude m_{odd} and with an adjacent-bar phase difference $\varphi = \pi/2$, detector output is proportional to the product of m_{odd} and m_{even}, regardless of the shape of the receptive fields. It can be shown that our linking assumption (see Section 1.A.3 below) implies that the probability of a correct direction response is a monotonic function of the product of m_{odd} and m_{even}. Experiments 2 and 3 test this multiplicative law.

(4) *Discrete approximations to moving sine waves.* A second class of displays consists of F-bar approximations to moving sine waves, where $F = 5$ in most of our experiments (=9 in Experiments 3 and 6). Hence

$$L_j(t) = L_{0j} + m \sin(2\pi\omega t - \vartheta_j). \qquad (20)$$

In this case, Eq. (17) becomes

$$y_4(\sin) = \beta_\omega \sin\delta_\omega m^2 \sum_{k=1}^{F-1} \sum_{j=1}^{F-k} \sin(\vartheta_{j+k} - \vartheta_j) A_{jk}. \qquad (21)$$

When, in addition, the between-bar asynchrony $\vartheta_{j+1} - \vartheta_j = \vartheta$ (const.), Eq. (21) can be further simplified to

$$y_4(\sin) = \beta_\omega \sin\delta_\omega m^2 \sum_{k=1}^{F-1} \sum_{j=1}^{F-k} \sin(k\varphi) A_{jk}. \qquad (22)$$

As we shall see in Experiment 4, this equation can be used to estimate the A_{jk}'s and allows us to make inferences about the shape of the input receptive fields on the basis of these estimates (Experiment 4).

3. Linking Hypothesis

Our experiments are psychophysical (the subject has to determine the direction of motion) and hence are assumed to involve a process by which responses from a large group of detectors are combined. We need a voting rule that specifies how a left–right decision is reached on the basis of discordant detector responses.

We assume that the probability of judging a display L as moving to the right, $P(\text{right} \mid L) = 1 - P(\text{left} \mid L)$, is a nondecreasing function of the detector outputs. More specifically, let outputs from detectors $D_1, \dots, D_M$ be denoted by $y_{4,1}, \dots, y_{4,M}$, and let V be a function that is nondecreasing in all of its M arguments. Then Prob(right $\mid L$) = $V[y_{4,1}(L), \dots, y_{4,m}(L)]$. If L and L' are two displays such that $y_{4,i}(L) \geq y_{4,i}(L')$ for all i, then Prob(right $\mid L$) $\geq$ Prob(right $\mid L'$). In addition, we assume that V is antisymmetric in the following sense: $V(z_1, \dots, z_M) = 1.0 - V(-z_1, \dots, -z_M)$. That is, the probability of responding left to a display L is equal to the probability of responding right to a display L' that causes all detector outputs to reverse in sign but stay equal in absolute magnitude. This assumption implies that a display that leads to zero output for all detectors causes chance performance (50% correct).

Our linking assumption is quite general and includes both the additive case, in which the response depends on the sum of the detector outputs, and the maximum case, in which the response depends on the maximum of all detector outputs (as is the case in threshold models).

The general voting rule was adequate to generate predictions for all experiments, except Experiment 3, for which a more specific voting rule was necessary (the rule, however, was still sufficiently general to include both the additive and the maximum case).[17]

4. The Original Reichardt Model and the Temporal Phase Line

As we anticipated in Section 1.A there is a direct link between a simplified version of the original Reichardt model and the temporal phase line. We explain why here. We have to make the simplifying assumption that detector inputs are derived only from adjacent bars, i.e., $A_{jk} = 1$ if $k = 1$ and zero otherwise. In this case, Eq. (17) becomes

$$y_4(L) = \sum_{j=1}^{F-1} m_j m_{j+1} p(\vartheta_{j+1} - \vartheta_j). \qquad (17')$$

Let ωn_0 be the temporal frequency and $T_0 = 1/(\omega n_0)$ the period of the dominant Fourier component of h, i.e., of the Fourier component that has the largest coefficient ε_n [Eq. (3)]. The temporal phase of the dominant F in bar B_j is $2\pi \omega n_0 \vartheta_j$.

The temporal phase line, which connects a luminance peak of the dominant Fourier component in bar B_j with the nearest luminance peak of the dominant Fourier component in bar B_{j+1} has the property that, in terms of Fig. 1, it descends from left to right if $\vartheta_{j+1} - \vartheta_j$ is between zero and $T_0/2$ and ascends from left to right if $\vartheta_{j+1} - \vartheta_j$ is between $-T_0/2$ and zero. Equation (17') implies that detector output is positive in the first case and negative in the second case. It follows that the sign of the detector output is the same as the sign of the slope of the temporal phase line. Hence we can predict the direction signaled by this simplified version of the original Reichardt model by looking at the slope of the temporal phase line. We included the temporal phase line in the representation of the displays for each experiment because this line gives the reader some intuition for the various predictions of the original and elaborated Reichardt models.

5. Terminology

We refer to the model as described originally by Reichardt,[10] and as described by us in Sections 1.A and 1.B, as the original Reichardt model. The model with our modifications to prevent aliasing and with an explicit linking hypothesis is called the elaborated Reichardt model. Note that the original Reichardt model describes only a single detector, whereas the elaborated Reichardt model describes an entire system of elaborated Reichardt detectors. When in some context both models are equivalent, we speak of the Reichardt model.

D. Alternative Models

We now turn to a discussion of alternative models for calculating motion. Several of these models were developed in artificial intelligence and machine perception and were originally not proposed as serious models of human motion perception. We include these models to illustrate different solutions to the problem of motion detection and, ultimately, to show how, by experiments, it is possible to discover which of these is tenable as a model of human perception.

All models discussed in this section use the familiar notion of frame, where frame i is defined as $L(x, t_i)$. It is useful to distinguish between global and local models. Global models analyze the entire frame or a significant fraction of a frame. Local models make use of a large number of units that each calculate the direction of motion within a small area. Outputs from these units have to be combined to infer the direction of motion of larger areas. The Reichardt model is a local model.

1. Global Matches

Spatial correlation analysis. One of the simplest ways to make use of frames is to find, for each pair of successive frames, the amount dx by which the second frame has to be shifted in order to maximize the product-moment correlation coefficient between the two frames. This analysis has been used for cloud tracking from satellite photographs.[18,19] A convenient graphic representation of this analysis is the cumulative plot of dx as a function of i, where i refers to the frame number. For a moving, rigid pattern, the cumulative plot graphs the location of a fixed point of the pattern as a function of time. We call this graph the motion path generated by the model. Some minor variations are possible.

First, if velocity is constant, one can calculate the straight motion path that has the highest mean correlation rather than the (not necessarily straight) motion path that has the highest correlation for each successive pair of frames. This might be advantageous in the presence of noise. Second, one can use other similarity measures than correlation. For example, one can use covariance or, as in the shift-and-subtract technique suggested by Anstis,[2] the sum of the absolute brightness differences.

Spatial phase analysis. Each frame can be decomposed into spatial Fourier components. The motion path consists of a graph of the spatial phase of the most prominent component as a function of time. This method has been used by Lo and his collaborators.[20] Also, Anstis and Rogers[21] suggest a spatial phase analysis model.

2. Local Models

Local brightness matching. Braddick,[1] who studied the perceived motion of random-dot patterns, discussed mechanisms that for a given location in frame i look within a small radius for the location in frame $i + 1$ that has the same brightness. Thus, for each location, the mechanism calculates the smallest displacement dx in which the equal-brightness location is found. These displacements would subsequently be combined by higher-level processes. In the case of moving random-dot patterns the task faced by these higher-level processes seems formidable, because it requires determining which dots in successive frames go together (i.e., solution of the correspondence problem[22]). However, when the between-frame displacement is smaller than the distance separating pairs of locations having equal brightness within a frame, no such ambiguity exists. To illustrate, when $L(x, t_i)$ is sinusoidal and is displaced by less than π in successive frames and if the radius within which the mechanisms look is less than π, the dx's are the same for all locations.

Spatiotemporal gradient matching. When a pattern moves in direction dx, the luminance change over time at location x_0 is the same as the luminance change when one moves within frame i from x_0 to $x_0 - dx$. Conversely, by matching up luminance changes across frames at one location with changes between different locations within one frame, one can obtain information about dx. Of course, for a given location x_0 there typically is more than one choice of dx that has matching change values. However, this problem vanishes as one approaches infinitesimal values of dx. In fact, it can be shown that the velocity is given directly by $-P_t(x_0, t_0)/P_x(x_0, t_0)$, where P_t and P_x are the partial derivatives of $L(x, t)$ with respect to x and t evaluated at location x_0 and time t_0. A two-dimensional, and substantially more complicated, version of this procedure was originally proposed by Limb and Murphy[23] and later refined by Fennema and Thompson.[24]

Spatial edge detection combined with temporal luminance change. Consider a frame $L(x, t_i)$ depicting a right edge, i.e., a frame in which luminance increases with x. The luminance at a given location increases when this edge moves to the left and decreases when it moves to the right. For a left edge the opposite holds. By combining knowledge about what type of edge is present with the local time course of the luminance of the image, one can infer the direction of motion. This idea, which is similar to gradient matching, was implemented in a quantized fashion by Marr and Ullman,[25] who proposed a system consisting of (local) detectors that work as

follows. Each detector consists of three subunits. The first subunit is a particular type of edge detector proposed earlier by Marr and Hildreth,[26] namely, a zero-crossing detector with binary (0 or 1) output. In our one-dimensional case, a zero crossing of frame $L(x, t_i)$ is defined as a point where the second derivative with respect to x crosses zero. This subunit comes in two versions, which are tuned to right and left edges. The second subunit consists of a linear temporal filter followed by a threshold device; it calculates the time derivative of the luminance in approximately the same patch of the visual field as processed by the edge-detecting subunit. This subunit also comes in two versions. Version T^+ has 1 as output when the time derivative is positive and 0 otherwise; vice versa for T^-. The third subunit of the detector performs an AND operation on the outputs of the first two subunits. Thus there are four types of detector, depending on the type of edge detector (right versus left) and the type of T unit. Two of these (left AND T^+, right AND T^-) are tuned to rightward motion; the other two (left AND T^-, right AND T^+) are tuned to leftward motion.

E. Current Psychophysical Evidence for the Original Reichardt Model

Applications of the original Reichardt model to human motion perception have been limited in number. Moreover, as we demonstrate below, these studies provided only weak evidence favoring the model over its competitors.

Schouten[27] studied apparent reversals in the perceived direction of a rotating, high-contrast grating at temporal frequencies ranging from 15 to 90 Hz. Schouten attributed this reversal to (in our terminology) spatial aliasing in the original Reichardt model. To explain why the reversal occurred at higher temporal frequencies only, Schouten would have to assume that, in the detector population, the distance $x_{right}-x_{left}$ increases with sensitivity to higher temporal frequencies (which may not be unreasonable; see Experiment 4, below). Another possibility is that the voting rule V involves thresholds, while, at the same time, at high temporal frequencies increasingly more detectors have temporal aliasing problems. That is, detectors with negative $\sin \delta_\omega$ have sufficiently low β_ω to stay below threshold at low or medium contrasts but not at high contrasts.

Schouten's reversal can also be explained by some of the alternative models. For example, the T detectors in the Marr and Ullman model[25] fail to function properly when the temporal period of the signal becomes too short. The same type of explanation could be given in terms of a neural implementation of the model by Fennema and Thompson,[24] because this model also requires calculation of a time derivative.

Foster[28] also used rotating radial gratings as displays, but only a small segment of the grating was visible through a window. Of primary relevance for the original Reichardt model is Foster's stationary stroboscopic effect, which is the perception of wavering motion when the window through which the rotating grating is viewed is larger than the width of a single white or black bar or, equivalently, larger than half of a spatial period. Foster explained this effect by positing that the spatial periods of his displays were sufficiently short so that they would cause spatial aliasing for most detectors. By making the window less than half of a spatial period wide,

detectors that otherwise would give erroneous responses are silenced, since at least one of their input channels is outside the window.

However, this explanation leaves open the question of why Schouten obtained well-defined motion without a window over a range of temporal frequencies including the frequencies used by Foster. An additional theoretical problem is that alternative explanations are possible. For example, we can analyze Foster's experiments in terms of the zero-crossing model of Marr and Ullman.[25] Opening up a window beyond half of a spatial period creates additional zero crossings that may affect perception.

There are also empirical problems with Foster's stationary stroboscopic effect. First, the task was a subjective judgment of well-defined motion, which, in fact, is difficult to define. Second, the upper bound on temporal frequency beyond which no well-defined motion could be seen was quite low (about 5 Hz), almost an order of magnitude below the temporal frequency at which direction can be discriminated. Third, we have been unable to replicate the phenomenon in our laboratory.

2. GENERAL METHODS

Except where noted, all seven experiments reported used the following methods.

A. Displays

Displays were produced on a computer-driven 0.30-m by 0.40-m Hewlett-Packard 1310A oscilloscope with a fast, white P4 phosphor; the displays were viewed binocularly with a natural pupil in a dimly lighted room. Displays consisted of adjacent, parallel vertical bars, each measuring 0.044 deg horizontally and 0.44 deg vertically. Except where noted, we used five bars without spacing. Thus the typical display measured 0.22 deg horizontally and 0.44 deg vertically.

We used a small display to minimize the effects of spatial inhomogeneity of the detector population. Preliminary experiments showed that as the distance between bars increased from zero to 0.22 deg, the strength of perceived motion fell to zero. Therefore 0.22 deg is the largest display width needed to study a detector of interest. The usage of five bars in most experiments was dictated by a compromise between two factors. Sperling[29] suggested that two bars would be the theoretically optimum display, but, in the present study, it proved much easier to collect data with multiple-bar displays (because they induced much stronger and less ambiguous perceived motion for reasons that are not yet entirely clear). The increased accuracy of performance with multiple-bar displays is much in excess of what can be predicted from probability summation of pairwise bar combinations. On the other hand, when the number of bars is substantially larger than five, the theoretical requirement of a particular phase between adjacent bars ($\pi/2$ in Experiments 2 and 3), in conjunction with the requirement that overall display size should be 0.22 deg or less, would produce extremely high spatial frequencies that are not suitable for these experiments. The most important reason for using five-bar displays is that certain tests in Experiment 4 (that compare $\varphi = \pi/4$ with $\varphi = 3/4\pi$) can be analyzed mathematically only for five-bar displays. Thus, although the apparatus was capable of producing many bar

gratings that closely approximated spatial sine waves, the experiments required the spatial bar patterns illustrated in Fig. 1.

The general mathematical form of the displays is given in Eq. (2). We refer to $h(t)$ as the modulation function of the display. In most cases, the L_{0j}'s and m_j's are the same; we denote their common values of L_0 and m, respectively. To simplify the exposition, we give luminances in terms of a reference luminance, L_{ref}, of 51 cd/m². Actual luminance is $L_{oj} \times L_{ref}$. By this convention, $L_0 = 1$ for most experiments.

The surrounding blank part of the CRT surface had a luminance of 0.7 cd/m² produced by two incandescent lamps. We put a black fixation mark (diameter: 2.5 min of visual angle) on the CRT surface in the center of the display.

B. Accuracy of Discrete Approximation to Temporal Sinusoidal Luminance Modulation

Displays with sinusoidal modulation functions consisted of 8-, 12-, or 24-point approximations (Fig. 1 contains 8-point approximations). These approximations were not always optimal, because only a small number of discrete luminance values could be used. However, Fourier analysis of the luminances revealed that, except at the smallest possible modulation amplitude (which was used only in one condition in Experiment 4), the fundamental frequency carried at least 90.5% of the total power. No individual harmonics carried more than 4%. For all modulation functions used, including sine approximations, we report the modulation of the computed Fourier components as a fraction of L_0 (not the difference between the maximum and minimum luminances of the luminance modulation function actually presented).

C. Trials

Except for the first trial of a block, in which the subject pushed a button to initiate the block, new displays were presented automatically at a (constant) time interval after termination of the preceding trial. The time interval varied across experiments from 0.9 to 1.2 sec. Each display lasted 1.32 sec. Leftward- and rightward-moving patterns occurred with equal frequency. The subject's task was to judge direction of movement. The judgment was made by pushing one of two buttons. We tabulated the percentage of correct (as *a priori* defined in each experiment) judgments of direction of motion. Because in several experiments the datum of interest was which, if any, direction the subject would perceive, no feedback could be given in these experiments. To make experiments consistent with each other, no feedback was used in the remaining experiments either. Pilot experiments showed that, after an initial training session without feedback, feedback does not further improve performance. We believe that this is partly a result of the fact that, except for Experiment 7, blocks contain several modulation depth levels mixed together, including suprathreshold levels. This enhances the subjects' capability to stay focused on relevant stimulus aspects.

Except for temporal frequency, experimental conditions were mixed within blocks. The reason that trials were blocked by temporal frequency is that pilot studies indicated that mixing temporal frequencies within blocks had a negative effect on performance. Sessions lasted approximately 1 h and always consisted of a practice block of 48 trials (45 in Exper-

iment 8) followed by 6 groups of 4 test blocks of the same length as the practice block. Intermissions between blocks were about 15 sec, intermissions between groups about 60 sec. Subjects were not given a dark-adaptation period. Occasional comparison of the first test blocks with the last test blocks revealed no systematic differences.

D. Subjects

Three subjects (two naive subjects, JP and NB, and the first author) served in the experiments. All subjects had at least 20:20 vision, one without correction, the others with correcting spectacles.

3. EXPERIMENTS

A. Experiment 1. Reichardt versus Korte

This experiment was performed early in our work in order to determine whether it was worthwhile to pursue Reichardt-type models. One of the most remarkable properties of the original Reichardt model is that the optimal between-bar asynchrony for motion detectors, $\vartheta_{j+1} - \vartheta_j$, depends on the temporal frequency composition of $h(t)$ [see Eqs. (4)–(8)]. To illustrate, suppose that detectors only look at adjacent bars (i.e., $A_{jk} = 1$ if $k = 1$, and 0 otherwise) and that $h(t)$ has only one prominent Fourier component (with temporal period T). Then the optimal asynchrony is simply $T/4$, regardless of what the value of T is. This follows from the term $\gamma_{right,n} - \gamma_{left,n}$ in Eq. (8a). Hence the optimal between-bar asynchrony varies with the temporal frequency of the most prominent component of $h(t)$. Obviously, this argument can be extended to any receptive-field arrangement. However, the precise form of the relation of optimal asynchrony to temporal frequency is determined by the receptive-field arrangement. For example, if detectors only look at bars that are two bars apart (i.e., $A_{jk} = 1$ if $k = 2$, and 0 otherwise), then the optimal asynchrony would be $T/8$.

In the classical studies of apparent motion, optimal asynchrony between bars in successive displays has been found to be determined primarily by interbar distance.[30] To the extent that our displays can be considered as direct generalizations of these two-bar two-view situations to an F-bar multiple-view situation, we would expect optimal asynchrony to be determined by the between-bar spacing. From Korte's second law, it follows that if we keep the spacing constant, optimal asynchrony would remain constant independently of the temporal frequency of the waveform.

We tested Korte's prediction that, for a fixed spatial configuration, the optimal asynchrony is independent of temporal-frequency content by comparing the effects of asynchrony for two modulation functions $h(t)$ and $h'(t)$. Function $h(t)$ is an ordinary 8-point approximation to a sinusoid (Figs. 1, 5A, and 5B) with period T. Function $h'(t)$ is a permutation of the same set of eight luminance values used to construct $h(t)$ (Figs. 5C and 5D). Fourier analysis of $h'(t)$ reveals that it consists primarily of a component having four times the temporal frequency of the fundamental of $h(t)$ and hence period $T' = T/4$. Because we constructed $h'(t)$ by permutation of $h(t)$, differences in performance between the two functions cannot be attributed to the set of luminance values used and hence not to differences in average luminance or
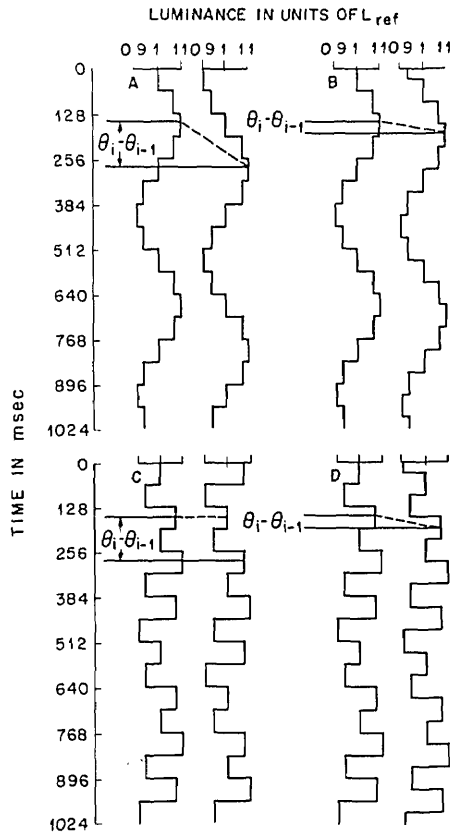
LUMINANCE IN UNITS OF L_{ref}

Fig. 5. Displays used in Experiment 1. Only modulations in the first two (of five) fields are depicted. A, same sinusoidal modulation function as in Fig. 1, 1.95 Hz, with asynchrony $\vartheta_{j+1} - \vartheta_j$ of 128 msec (hence phase difference φ of $\pi/2$ rad). B, same sinusoidal modulation function as in Fig. 1, but with asynchrony of 32 msec ($\varphi = \pi/8$ rad). C, modulation function that is a permutation of the function used in A and B. Between-field asynchrony is 128 msec, but the dominant fourth harmonics are in phase and hence convey no directional information. D, same modulation function as in C, but asynchrony is 32 msec ($\pi/2$ rad in terms of the dominant fourth harmonic).

contrast or any other statistic defined on this set. Of course, the Reichardt model states that the order in which the luminance values are presented is critical because it determines the temporal-frequency content.

We used two asynchronies. The first asynchrony was $T/4$ (Fig. 5A). According to the original Reichardt model, with the added assumption that detectors look at adjacent bars only, the asynchrony of $T/4$ is optimal for discrimination of the direction of $h(t)$. This asynchrony causes the largest Fourier component of $h'(t)$, which has period $T/4$, to have zero between-bar phase differences. In terms of Eq. (8a), $\sin(\gamma_{right,4} - \gamma_{left,4}) = 0$ regardless of whether detectors look at adjacent or at nonadjacent bars. Hence the original Reichardt model predicts that for $h'(t)$, performance should be at chance levels. The second asynchrony was one fourth the duration of the first asynchrony. Now, the phase differences are $T/16$ for $h(t)$ (Fig. 5B) and $T'/4$ for $h'(t)$ (Fig. 5D). Again assuming that detectors look only at adjacent bars, the original Reichardt model predicts that performance on $h(t)$ should decrease relative to the first asynchrony [because $\sin(\pi/8)$ is much less than $\sin(\pi/2)$], whereas performance on $h'(t)$ should increase [because $\sin(\pi/2)$ is much larger than $\sin(0) = 0$].

In order to make predictions for the elaborated Reichardt model, we first have to find the primary spatiotemporal sine-wave components of these displays. If each of the four types of displays is dominated by one sine wave, we can, on the basis of our assumption of absence of spatial and temporal aliasing, predict the direction of perceived movement, which will simply be the movement of the dominant sine wave. It turns out that, ignoring stationary sine waves (which, as we stated earlier, have no effect on detector response), the displays are indeed each dominated by one sine wave. At the asynchronies of $T/4$ and $T/16$, these dominant components are, for $h(t)$, $\sin(2\pi n\omega t + 2\pi 5.68x)$ and $\sin(2\pi n\omega t + 2\pi 1.42x)$, respectively, and, for $h'(t)$, $\sin(2\pi n\omega t)$ (homogeneous flicker) and $\sin(2\pi n\omega t + 2\pi 5.68x)$, respectively. Which asynchrony leads to better performance for $h(t)$ cannot be predicted, because this depends on the receptive-field shape. As for $h'(t)$, it can be predicted that performance on $\sin(2\pi n\omega t)$ should be at chance, because detector response to homogeneous flicker is zero no matter what the receptive-field arrangement is [in Eq. (17), $\vartheta_{j+k} = \vartheta_j$, making the factor $p(\vartheta_{j+k} - \vartheta_j)$ zero].

1. Method

Modulation functions $h(t)$ and $h'(t)$ are as described above and in Fig. 5. We used a temporal period T of 512 msec (1.95 Hz). Between-bar asynchronies were 32 and 128 msec. Sine-wave modulation was 0.14 for $h(t)$, and the same luminances were used for $h'(t)$. Two subjects, NB and JvS, made 288 observations in each of the four (h versus h', 32- versus 128-msec asynchrony) experimental conditions.

2. Results

For subjects NB and JvS, performance on modulation pattern h was 77 and 88%, respectively, at 128-msec asynchrony and 56 and 54% correct at 32-msec asynchrony; performance on modulation pattern h' was 49 and 49% correct at 128-msec asynchrony and 94 and 94% correct at 32-msec asynchrony. These results indicate that optimal asynchrony depends on temporal-frequency content of the modulation function. Moreover, the results are precisely what one would predict on the basis of the original Reichardt model with the added assumption that detectors look only at adjacent bars. Obviously, these results are completely inconsistent with Korte's idea of an optimal between-bar asynchrony that is independent of temporal frequency.

B. Experiment 2. Pulse Reversal

The multiplicative law [Eq. (19)] states that when bars that are an even number of bars apart are either in phase (zero asynchrony) or in counterphase (π asynchrony) with each other and when odd-numbered bars have amplitude m_{odd} and even-numbered bars have amplitude m_{even}, then performance depends on the product of m_{odd} and m_{even}. In this experiment, we test Eq. (19) by reversing the sign of m_{odd}, i.e., we invert the modulation of odd-numbered bars. Both the original and the elaborated Reichardt models predict that this should lead to perception of motion in the reverse direction, regardless of what the shape of h is. Since neither h nor the absolute magnitude of the product $m_{odd}m_{even}$ changes when we reverse the sign of m_{odd}, both models predict perfect reversal of perceived motion.
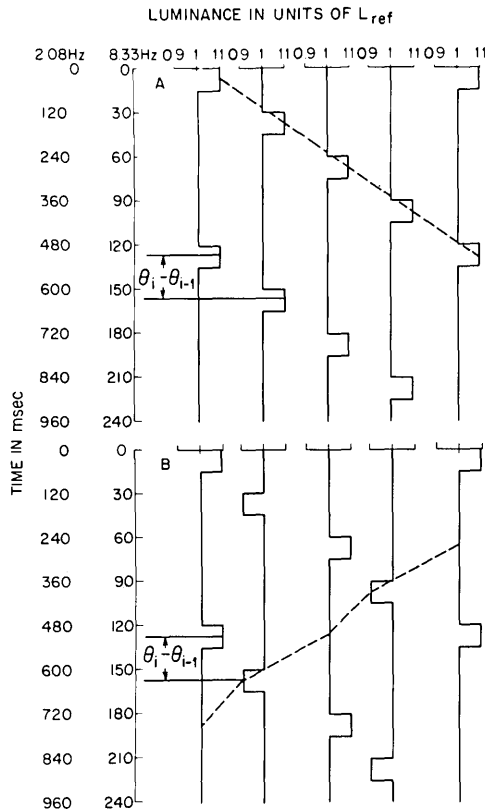
Fig. 6. Display used in Experiment 2. A, modulation function consists of pulses. Phase difference is 120 msec for 2.08-Hz display; 30 msec for 8.33-Hz display (both $\pi/2$ rad). B, same as in A, except that modulation functions in even fields are reversed in sign. Note that in B, fields containing same-sign pulses are either in phase or in counterphase and thus do not contain direction information.

In this experiment, we chose h to be a pulse function rather than a sinusoid (Fig. 6). This means that we have to take into account higher temporal harmonics of h. When $m_{odd} > 0$, the fundamental temporal frequency ω has between-bar phase differences of $\pi/2$; the second harmonic (2ω) is in counterphase across bars ($\pm\pi$), whereas the third harmonic (3ω) has a between-bar phase difference of $-\pi/2$ (or $+3\pi/2$). In general, the kth harmonic has a phase difference of $k\pi/2$. Of course, when $m_{odd} < 0$, the kth harmonic has a phase difference of $-k\pi/2$. Thus, regardless of the sign of m_{odd}, the third, seventh, etc. harmonics indicate a direction of movement opposite that indicated by the fundamental, whereas the even-numbered harmonics indicate no direction because their between-bar phase differences are multiples of π. It follows from the property of segregation of temporal frequencies that performance (defined as the proportion of direction responses consistent with the direction of the fundamental) should decrease when the temporal period of the display becomes sufficiently long to allow the harmonics to fall inside the range of optimal temporal-frequency sensitivity. (Later, in Experiment 4, we show this range to extend to approximately 15 Hz.)

The spatiotemporal sine-wave components of the display indicate the same pattern of directions of higher harmonics. That is, even harmonics consist of either counterphase gratings or homogeneous flicker, whereas odd harmonics have directions that are either opposite or the same as the direction

of the fundamental sine wave. By the assumption of nonaliasing, the elaborated Reichardt model thus makes the same prediction concerning the effects of harmonics as the original model.

It should be noted that, in the context of pulse stimuli, the prediction of perfect reversal of perceived motion is particularly strong, because the temporal modulation patterns in adjacent bars are much more similar to each other in the standard display (Fig. 6A) than in the reversal display (Fig. 6B).

1. Method

We used as modulation function $h(t)$ a periodic pulse function having a value of 1 in the initial 1/8 of a period and zero elsewhere. Fourier series analysis of $h(t)$ revealed that the powers of the harmonics relative to the power of the fundamental slowly declined and reached 50% at the tenth harmonic.

We compared three conditions [$(m_{even}, m_{odd}) = (m, m), (m, -m), (-m, -m)$] corresponding to whether the m_j's were (1) all positive (Fig. 6A), (2) positive and negative in adjacent bars (Fig. 6B), or (3) all negative (not illustrated). We used two temporal periods of $h(t)$ [120 and 480 msec (8.33 and 2.08 Hz)] and two modulation levels (0.062 and 0.099 for NB, 0.037 and 0.049 for JvS). In each of these 12 experimental conditions (3 m_j combinations, 2 temporal frequencies, and 2 modulation levels), 192 observations were made.

2. Results

The data from the 2.08-Hz condition are not graphed since neither subject was able to see a consistent direction of motion, thus resulting in chance performance. This is astounding because, in the (m, m) and $(-m, -m)$ conditions, the display consists of an increased luminance that simply passes from one bar to the next. According to the Reichardt model, we are not aware of this simple state of affairs, because, by the property of segregation of temporal frequencies, we analyze this display into sine waves that move in conflicting directions. Figure 7 contains the results for the 8.33-Hz condition. It shows that perceived direction of motion is completely predicted by the fundamental component. Both subjects show near-perfect reversal of perceived motion for the reversed-pulse stimuli.

C. Experiment 3. Multiplication of Alternate Bar Amplitudes

Experiment 3 constitutes an additional test of the multiplicative law [Eq. (19)]. Experiment 3 focuses on the monotonicity property of multiplication: An implication of the multiplicative effect of m_{odd} and m_{even} on accuracy of motion detection is that, when we hold m_{odd} fixed, performance should increase monotonically as we increase m_{even}, and vice versa. This property is counterintuitive for two reasons. First, it violates the intuition that perception of motion is enhanced by between-bar similarity. More specifically, if we start out with m_{odd} and m_{even} at some low value and then increase one of the amplitudes, this leads to a decrease in the similarity of the modulation functions in adjacent bars. For example, similarity as measured by the mean-squared difference between bars is reduced by unequal amplitudes of modulation. Similarity as measured by the product-moment correlation coefficient is neither reduced nor increased by
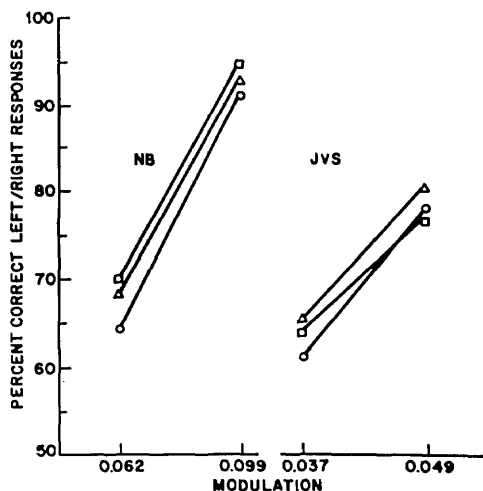
Fig. 7. Results of Experiment 2. The ordinate indicates the proportion of direction responses that are consistent with the objective direction of the dominant temporal-frequency component. The abscissa indicates modulation depth. Displays with sign-reversed pulses in even fields (△) yield performance roughly equivalent to average performance with positive pulses (□) and negative pulses (○).
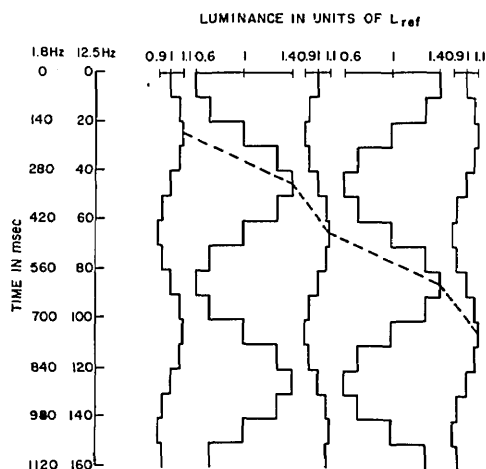

Fig. 8. Display used in Experiment 3. Modulation function is sinusoidal with asynchrony of 140 msec for 1.8-Hz display ($\pi/2$ rad) and 20 msec for 12.5-Hz display ($\pi/2$ rad). All fields have the same average luminance, but even-field modulation is four times larger than odd-field modulation. Note that fields with large modulations are in counterphase with each other and hence do not contain directional information.

unequal modulation. Second, suppose that the modulations are sufficiently low in all bars to cause chance-level performance. Suppose that we now increase the modulations of the odd-numbered bars. In effect, we are adding a display that by itself does not contain direction information. That is, increasing m_{odd} from m' to m'' is equivalent to adding the function $L'(x, t)$, which has zero luminance for even-numbered bars and is equal to $(m'' - m')L_j(t)$ for odd-numbered bars. Thus $L'(x, t)$ consists of alternate bars that are either in phase or in counterphase with each other and hence does not contain directional information. Nevertheless, the Reichardt model predicts that adding this particular ambiguous display will strengthen whatever perception of motion the original display might have evoked. This monotonicity prediction is here tested in a wide range of conditions.

1. *Method*

The displays consisted of modulation functions of the type $h(t) = \sin(2\pi n\omega t - j\pi/2)$, where $j = 1, \ldots, 9$. Figure 8 illustrates five of the nine bars, with $m_{odd} = 4m_{even}$. The amplitudes m_{odd} and m_{even} were manipulated factorially: m_{odd}/L_0 took on values of 0.023, 0.042, 0.064, and 0.085 and m_{even}/L_0 took on values of 0.023, 0.042, 0.085 and 0.229. It should be noted that, when both m_{odd}/L_0 and m_{even}/L_0 are at 0.023, performance is at chance levels. Two temporal frequencies were used (1.8 and 12.5 Hz). Two subjects, NB and JvS, each made 72 observations in each of the 32 experimental conditions (four values of m_{odd} and m_{even} and two temporal frequencies).

2. *Results*

Data for the 32 conditions are shown in Fig. 9. In Fig. 10, we replotted the data as a function of the product of m_{odd} and m_{even}. Figure 9 clearly shows that the data satisfy monotonicity. That is, except for minor statistical fluctuations, whenever either m_{odd} or m_{even} is fixed and the other is increased, performance increases or, when it is already almost perfect, does not decrease. In addition, although performance was at chance levels when both m_{even}/L_0 and m_{odd}/L_0 were at
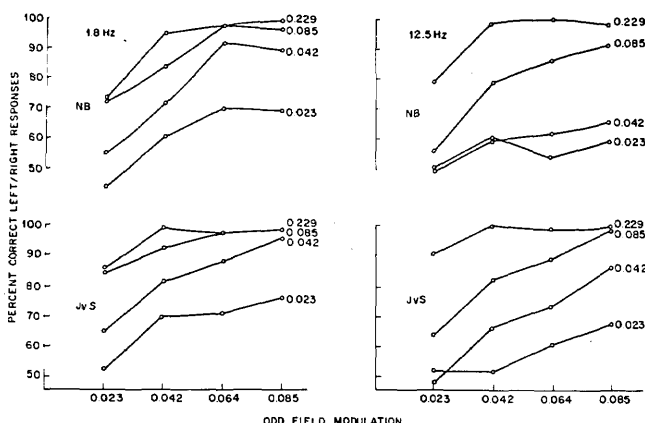

Fig. 9. Results from Experiment 3, the effects of odd- and even-field modulation. The abscissa is m_{odd}; the curve parameter is m_{even}; the ordinate is the percent of directionally correct responses. Separate panels are shown for 1.8 and 12.5 Hz for both subjects, NB and JvS.
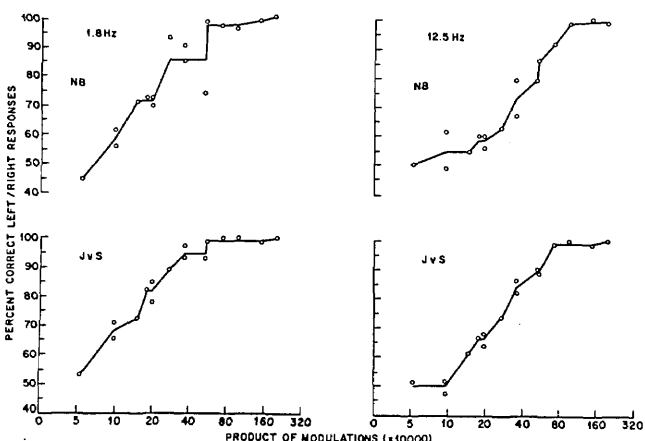

Fig. 10. Results from Experiment 3. Percent correct left–right responses as a function of the product of m_{odd} and m_{even}. (○), Data points; solid curve indicates isotonic regression.

0.023, performance increased dramatically when either of the two amplitudes was increased. That is, adding an ambiguous stimulus to an unambiguous threshold stimulus brought the combination above threshold of unambiguously correct responses.

The multiplicative law can be tested directly by comparing displays that have equal products of m_{odd} and m_{even} achieved in different ways. For example, the following (m_{odd}, m_{even}) pairs have the same product and induce the same levels of performance: (0.23, 0.85), (0.43, 0.43), (0.85, 0.23). The relationship between the product $m_{odd}m_{even}$ and performance is illustrated in Fig. 10. This relation varies somewhat depending on the subject and the temporal frequency but obviously is a good predictor of performance over an enormous range of 48:1 in our data. Other combinations of m_{odd} and m_{even} (such as summing m_{odd} and m_{even} to predict performance) fail miserably. For example, summation predicts that performance with (m_{odd}, m_{even}) = (0.085, 0.085) should be worse than with (0.023, 0.229), whereas our multiplicative law predicts the opposite. Clearly, the data show the latter. We conclude that the data from Experiments 2 and 3 provide strong support for the multiplicative law.

D. Experiment 4. Temporal and Spatial Frequency

In this experiment, we measure the effects of temporal frequency (ω) and phase difference (φ) on the detection of motion that is defined by sinusoidal modulation functions $h(t)$ = $\sin(2\pi n\omega t - j\varphi)$ [cf. Eqs. (2) and (20)]. This experiment serves several purposes. First, these displays can be considered as being approximations to rigid, moving sine waves. The temporal frequency (in hertz) is ω; the spatial frequency f [in cycles per degree (cpd)] is 3.62φ; the velocity (in degrees per second) is $\omega/(3.62\varphi)$. Hence we can verify whether the data obtained by our methods are consistent with standard data on the human spatiotemporal frequency response obtained with spatially finer approximations of sine waves than our displays, in particular, the data obtained by Kelly[31] and Burr and Ross.[32]

Second, this experiment allows for a direct test of the point input assumption of the original Reichardt model. The test consists of comparing $\varphi = \pi/4$ with $\varphi = 3\pi/4$ in a display that has five bars ($F = 5$). [Refer to Eq. (22), which describes the output of a detector for these inputs.] For the distance between bars, $k = 0, 1, 3, 4$, $\sin(k\pi/4) = \sin(k3\pi/4)$, whereas for $k = 2$, $\sin(k\pi/4) = 1 = -\sin(k3\pi/4)$. Hence, unless there is temporal aliasing (i.e., $\beta_\omega \sin \delta_\omega < 0$), the detector output y_4 can never be larger for $\varphi = 3\pi/4$ than for $\varphi = \pi/4$, unless $\Sigma_{j=1}^3 A_{j2} < 0$. As we pointed out earlier, for the original Reinhardt model, A_{jk} is either 0 or 1, so that $\Sigma_{j=1}^3 A_{j2} \geq 0$. It follows that the original Reichardt model predicts that, under all circumstances except temporal aliasing, performance with $\varphi = \pi/4$ should be at least as good as performance with $\varphi = 3\pi/4$.

Third, the experiment gives information about the particular form of the input receptive fields $r_H(x)$. Again, the comparison between the $\pi/4$ and $3\pi/4$ phase differences is critical. As stated above, performance with $3\pi/4$ can exceed performance with $\pi/4$ only when $\Sigma_{j=1}^3 A_{j2}$ is negative. Computer simulations[33] showed that this sum is nonnegative for a wide range of receptive-field shapes that are single-peaked, symmetric, and identical except for location. This was the case regardless of the location of the receptive fields relative

to the display center and the distance between the receptive-field centers. Further simulations showed that minor violations of these three properties do not cause the sum to be negative, provided that receptive fields cross at only one location. On the other hand, the sum is negative for the triple-lobed receptive-field pairs depicted in Fig. 3, provided that the detectors are centered on the display and that the spatial frequency of the display with $3\pi/4$ phase difference is f_0 [Eqs. (12) and (14)]. In other words, the sum is negative for precisely those detectors that respond most strongly to the $\varphi = 3\pi/4$ display, because they are centered on the display and have a spatial frequency response that peaks close to the spatial frequency of the display. Although we do not have a rigorous mathematical proof, these considerations strongly suggest that $\Sigma_{j=1}^3 A_{j2}$ is negative only for receptive-field pairs that cross more than once and that must therefore have several on and off areas. If we were to find that performance for $\varphi = 3\pi/4$ exceeds that for $\varphi = \pi/4$ over a range of temporal frequencies (thereby excluding temporal aliasing), we would regard this as strongly suggestive evidence for multilobed receptive fields. One may remark here that evidence for multilobed receptive fields may be obtained in the following, much simpler, way. Since $\varphi = 3\pi/4$ corresponds to a higher spatial frequency than $\varphi = \pi/4$, in which we find better performance in the former than in the latter condition, one could simply say that elaborated Reichardt detectors do not have a low-pass spatial-frequency response but, rather, a nonmonotonic spatial-frequency response such as a band-pass response. Because multilobed fields have a band-pass spatial-frequency response and single-lobed fields tend to have a low-pass spatial-frequency response, it would seem to follow that band-pass data reject single-lobed receptive fields. However, the factor $D(f)$ in Eq. (15) shows this argument to be incorrect: even when P_{left} and P_{right} are strictly decreasing (i.e., receptive fields are low-pass), the spatial-frequency response of the detector itself is always band-pass, since $D(0)$ = 0.

1. Method

The stimuli consisted of five bars with luminance function $h(t) = \sin(2\pi n\omega t - j\varphi)$. All combinations of three values of $\varphi(\pi/4, \pi/2, \text{and } 3\pi/4)$ and eight values of ω (1.04, 2.08, 4.17, 6.94, 10.42, 15.63, 20.83, and 31.25 Hz) were tested. The spatial frequencies corresponding to the three values of φ are 2.84, 5.68, and 8.52 cpd.

Seventy-five-percent-correct direction thresholds were measured with the method of constant stimuli. Several methods of interpolation were applied to the psychometric functions, but these all gave essentially the same results. We report thresholds obtained with isotonic regression analysis.[34]

Threshold determinations were based on 384–864 observations for each of the 24 experimental conditions.

2. Results

Seventy-five-percent thresholds as a function of temporal frequency are given in Fig. 11. Separate temporal modulation transfer functions (MTF's) are plotted for the three values of φ, $\pi/4$, $\pi/2$, and $3\pi/4$, which correspond to spatial frequencies of 2.84, 5.68, and 8.52 cpd. We see that for each value of φ, sensitivity in an inverted-U-shaped function of temporal frequency that reaches an optimum between 4 and 8 Hz. Sensitivity decreases sharply beyond 15 Hz. The MTF's
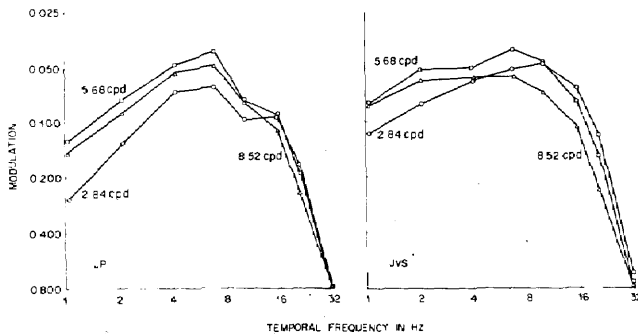
Fig. 11. Seventy-five-percent threshold modulations (m/L_0) for displays with sinusoidal modulation functions, as a function of temporal frequency. Phase differences are 0.25π rad (O), 0.5π rad (□), or 0.75π rad (△), corresponding to spatial frequencies of 2.84 (O), 5.68 (□), and 8.52 (△) cpd.

cross: at low temporal frequencies, sensitivity is highest for $\varphi = \pi/2$ and lowest for $\varphi = \pi/4$; at high temporal frequencies, sensitivity is highest for $\varphi = \pi/4$ and lowest for $\varphi = 3\pi/4$.

We now consider the issue of whether these data are consistent with existing data, in particular those of Kelly[31] and of Burr and Ross.[32] First, our modulation amplitude at threshold in the optimal spatiotemporal frequency range is about 0.04. This is a larger modulation amplitude than Kelly measured (0.005). However, Kelly used a stimulus area (about 44 deg^2) that was several hundred times as large as ours (about 0.1 deg^2). Our amplitude also is larger than the smallest amplitude measured by Burr and Ross. However, their viewing area was even larger than Kelly's (up to 12,100 deg^2).

Second, the overall shape of the temporal MTF at our range of spatial frequencies was similar to the shapes found both by Kelly and by Burr and Ross. The optimum is in the 4–8-Hz range, and sensitivity sharply declines for temporal frequencies over 15 Hz.

Third, in our data, as in the data of Kelly and of Burr and Ross, optimal spatial frequency depends on temporal frequency. In fact, our MTF's in Fig. 11 exhibit a crossover pattern—an irrefutable interaction. The MTF's of Kelly and of Burr and Ross exhibit a similar interaction. However, these authors find that the sensitivity to 8.52 cpd, while increasing relative to 2.84 cpd as temporal frequency goes down, never surpasses the latter. In other words, our data are different in that the 8.52-cpd MTF is elevated as a whole relative to the 2.84-cpd MTF. Two differences in procedure could explain the difference in vertical displacement of our temporal MTF's and those of Kelly. First, we showed many fewer spatial cycles of sine waves. For example, for the $\pi/4$ display (2.84 cpd), we showed only half a cycle. Second, our presentation was exclusively foveal. Both factors bias our results in favor of patterns having high spatial frequencies, causing high-spatial-frequency MTF's to rise to an altitude in Fig. 11 that enables them to intersect the other curves.

In summary, where comparable, our data, obtained with discrete approximations to moving spatial sine patterns, are quite consistent with data obtained by Kelly and by Burr and Ross with continuous spatial sines. Our thresholds are somewhat higher than theirs, especially at low spatial frequencies, because of our much smaller displays.

The superior performance at $\varphi = 3\pi/4$ over $\varphi = \pi/4$ at low

temporal frequencies shows that the original Reichardt model with point input assumption is incorrect. The obvious correction is the assumption of extended receptive fields. Our computer simulations suggest that the fields must have off areas in addition to on areas and that the fields must cross at several locations. Both receptive-field pairs in Fig. 3 have these properties.

The crossovers in our MTF's (and the corresponding interactions in the MTF's of Kelly and of Burr and Ross) ultimately require an additional complication of the Reichardt model, i.e., the assumption of an inverse relationship between optimal spatial and temporal frequency. The most obvious way to implement such a relationship is by assuming that there are at least two kinds of detectors; specifically, that receptive fields are smaller for fast than for slow detectors (here, fast versus slow is defined in terms of the temporal frequency ω that maximizes $\beta_\omega \sin\delta_\omega$).

E. Experiment 5. Unequal Bar Mean Luminances

As noted before, in both the original and the elaborated Reichardt models, detector output is independent of $L_0(x, t)$, i.e., of the stationary component of $L(x, t)$. In this experiment, we test this property by investigating the effects of adding a constant amount of background luminance to alternate bars of a motion display (i.e., a display with sinusoidal modulation functions with modulation m that represent motion). Thus, in one condition (uniform background), $L_{0j} = L_0$, for all j. In the other condition (grating background, Fig. 12), $L_{0j} = L_0$ if j is even and $L_0 + c$ if j is odd. To illustrate the prediction of no effect of grating background, note that the temporal phase line in Fig. 12 does not deviate from its counterpart in Fig. 1, which depicts the uniform-background condition. The comparison of uniform and grating backgrounds yields important information, because, as we shall see, all but one
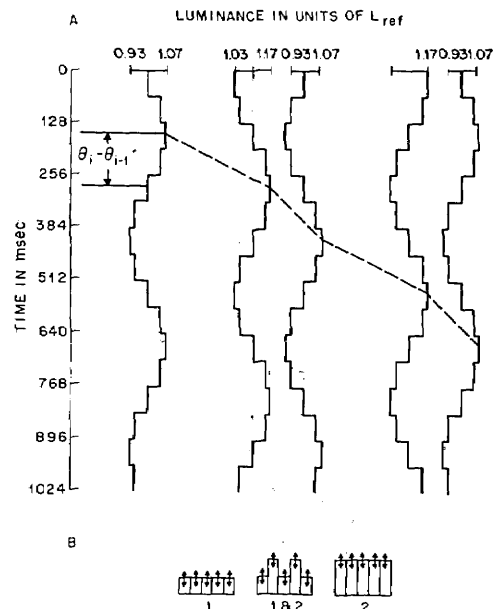

Fig. 12. One of the three types of display (1&2) used in Experiment 5. Modulation function is sinusoidal, with same modulation m_j but different average luminance L_{0j} across fields. Other displays (1) and (2) in this experiment have the same average luminances across fields. Inset: Alternative representation of displays (1), (1&2), and (2) used in Experiment 5. Arrows indicate modulation m_j, and height of vertical bars indicates average luminance L_{0j}.

(spatial phase analysis) of the alternative models predict that the grating background should have detrimental effects on performance.

1. Precautions

(1) Eye movements. There is a risk that the data may fail to fulfill the prediction of no effect of grating backgrounds for the wrong reasons. There is the possible role of eye movements. The stationary grating background increases or reduces the response of motion detectors to the motion stimulus, depending on whether the eye movement is against or with the stimulus movement. These with/against effects need not cancel each other; in fact, most considerations give the upper hand to sensitivity reduction over increase. Kelly[31] found that the detection threshold for a moving sine wave is increased by about 10% when a stationary sine wave that has the same amplitude and spatial frequency as the moving sine wave is added. Kelly attributed this small effect to uncontrolled eye movements. Since we do not use stabilized images, the same possibility exists in our experiments. However, in our procedure, display duration is relatively short (1.32 sec) as compared with that of Kelly, whose method of adjustment probably required a much longer duration. Moreover, even when the method of adjustment is used, it has been shown[35] that well-trained observers viewing moving sine waves have negligible eye movements, provided that a fixation mark is used. Thus it is unlikely that eye movements significantly affect our data.

(2) Adaptation. The predictions of the Reichardt model are based on the assumption that none of the components changes its characteristics with changes in mean luminance. Clearly, the human visual system changes enormously with adaptation level, a complication we strive to avoid, for the time being, in the expectation that the mechanisms of adaptation and of motion detection can be treated separately. We expect to observe that the motion threshold is independent of L_0 only for small changes in $L_0(x, t)$, for which the assumption of linearity of the SF and TF components is reasonable accurate. Because of adaptation, the prediction of no effect of grating background on motion thresholds must be tested by comparing a grating background with a uniform background of the same average luminance. To obtain sensitive bounds on this prediction, we actually use two uniform backgrounds as our controls: (1) $L_{0j} = L_0$, and (2) $L_{0j} = L_0 + c$. The mean background luminance in the grating condition lies between backgrounds (1) and (2); our prediction is that the motion thresholds against grating backgrounds must lie between thresholds against uniform backgrounds (1) and (2).

2. Predictions of Alternative Models

The spatial phase analysis model[20,21] is immune to addition of a stationary grating when the spatial frequency of the stationary grating differs from that of the moving grating, as is the case in the present experiment (see below). The reason is that the motion path for the moving sine wave is unaffected by addition of the stationary grating, the only effect of which is to produce its own motion path in the form of a horizontal line (indicating absence of motion). However, the remaining models predict detrimental effects of adding a stationary grating.

(1) Spatial correlation analysis.[2,18,19] Adding a stationary grating changes the spatial pattern in each frame; we cannot match up frames exactly anymore, because either the moving pattern or the stationary pattern will cause a mismatch. Computer analysis of our displays showed that, in particular, when $c \geq 2m$, no direction of motion can be inferred for any of the three similarity measures (i.e., product-moment correlation, covariance, and sum of absolute brightness differences). The path that maximizes the frame-to-frame similarity (by any measure) forms a zigzag pattern without obvious direction; the straight path that maximizes the mean frame-to-frame similarity is a horizontal line (no motion).

(2) Local brightness matching.[1] If this model sets a premium on matching points that are identical or nearly identical in luminance, then the situation where $c = 2m$ is particularly interesting. As can be demonstrated graphically, the only points of equal luminance in adjacent bars in successive frames are the peak in one bar and the trough in the other bar, but these two points indicate motion in the reverse direction.

(3) Marr–Ullman.[25] Nonmoving zero crossings are introduced in the added-grating stimulus, whereas there are none in the control displays, and up to one half of the moving zero crossings in the control conditions are eliminated or reversed in the added-grating condition. Thus motion detection that depends on zero crossings would be severely impaired by a grating background.

(4) Spatiotemporal gradient matching.[23,24] The spatial derivative at a fixed point on the moving sinusoid is affected by the value of the derivative of the stationary pattern at that location, whereas the temporal derivatives are unaffected. Hence the process that matches spatial and temporal derivatives is perturbed, and motion detection is also.

3. Method

We used the same sinusoidal modulation functions as in Experiment 4 to produce a discrete approximation to a rigidly moving sinusoidal grating. The critical variable was whether we added a constant luminance c of 5.9 cd/m² to L_{0j} for odd-numbered bars (Fig. 12). To control for effects of overall luminance, we compared three conditions: (1) $L_{0j} = 51$ cd/m² for all j; (1&2) $L_{0j} = 51$ cd/m² if j is even, 56.9 cd/m² if j is odd; and (2) $L_{0j} = 56.9$ cd/m² for all j. The remaining factors were ω (1.95, 7.81, and 15.6 Hz), φ ($\pi/4$, $\pi/2$, and $3\pi/4$), and m (at levels needed to ensure motion detection between 65 and 85% correct). For subject JP, $2m$ was larger than c in some conditions; for subject JvS, $2m$ was always smaller than c. Subject JP received at least 640 trials in each of the 27 [(1) versus (1&2) versus (2), ω, and φ] experimental conditions; subject JvS received 384 trials. Presentation was blocked by φ and, as usual, by ω.

4. Results

Overall, performance on displays of type (1&2) was between performance in displays (1) and (2): Performance in the three display conditions was 67.7, 65.6, and 64.9% correct for subject JP and 78.1, 74.9, and 72.9% for subject JvS. There was no trend consistent across both subjects that was related to either spatial or temporal frequency or that could be attributed to whether (for subject JP) c was larger or smaller than $2m$.

These data warrant two conclusions. First, a uniform background, which lowers stimulus contrast, produces a small but measurable sensitivity loss (2.8 and 5.2% for subjects JP and JvS, respectively). This indicates that direction discrimination mechanisms are not fully linear (independent of stationary uniform backgrounds); this does not come as a surprise. Second, and more important, adding a stationary grating has remarkably little effect beyond what can be explained by contrast reduction. We found this lack of effect of background grating over a large range of spatial and temporal motion frequencies.

F. Experiment 6. Adding a Stationary Sine Wave with the Same Spatial Frequency as the Moving Sine Wave

This experiment was a specific test of the spatial phase analysis model.[20,21] In the previous experiment, the spatial frequency of the stationary pattern (11.36 cpd) was always different from the spatial frequency of the moving sine wave (2.84, 5.68, and 8.52 cpd). The spatial phase analysis model was not rejected by this experiment, because (spatial) Fourier analysis keeps sine waves of different spatial frequencies separate. However, when the stationary and the moving sine wave have the same spatial frequency, something different happens (Fig. 13).

Within a frame, the moving and stationary sine waves combine to form a new spatial sine wave with the same spatial frequency as the component sine waves. However, the amplitude of this sine wave changes from frame to frame, and, more importantly, the spatial phase does not follow the same path as the spatial phase of the moving sine wave. In fact, the spatial phase path is a periodic function that approaches a straight horizontal line (no motion) as the amplitude of the stationary sine wave approaches infinity. In particular, when the stationary and moving sine waves have equal amplitudes, the spatial phase path moves back and forth over a range of half of a spatial cycle.

In the present experiment, we investigate the effects of adding a stationary sine wave to a moving wave; both have the same amplitude and spatial frequency. It should be noted that, in this experiment, zero crossings, as defined by Marr and Ullman,[25] follow the same back-and-forth path as spatial phase.
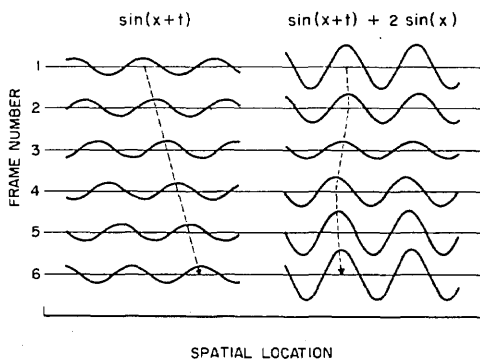

Fig. 13. Effects on the spatial phase path of a moving sine wave when a stationary sine wave with the same spatial frequency is added. Luminance of a given location in a frame is indicated by the vertical deviation from the horizontal dashed null line for the frame. Vertical dashed lines connect locations having the same spatial phase in successive frames. Left-hand panel: six successive frames of $\sin(x + t)$, where x denotes location and $t(=t_1, t_2, \ldots, t_6)$ time. Right-hand panel: six successive frames of $\sin(x + t) + 2\sin(x)$.
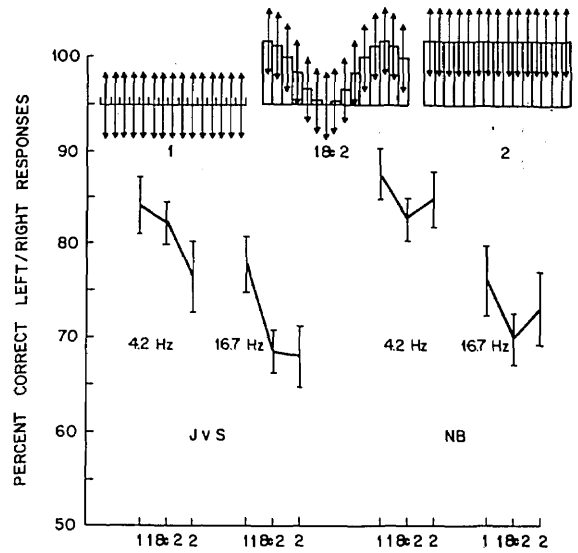

Fig. 14. Results of Experiment 6. Effects of adding a stationary sinusoid having the same spatial frequency ω and modulation m_j as a moving sinusoid. Error bars indicate 95% confidence intervals. Inset: Alternative representation of displays (1), (1&2), and (2) used in Experiment 6.

1. Method

The same procedure followed in Experiment 5 was used here with the following exceptions (Fig. 14, inset). First, conditions (1), (1&2), and (2) were defined as follows. In condition (1), $L_j(t) = L_0 - m + m\sin(2\pi\omega t - j\pi/6)$; in condition (1&2), $L_j(t) = L_0 + m\sin(2\pi\omega t - j\pi/6) + m\sin(-j\pi/6)$; and in condition (2), $L_j(t) = L_0 + m + m\sin(2\pi\omega t - j\pi/6)$. Thus the time-averaged luminance was $L_0 - m$ in all bars in condition (1), $L_0 + m$ in all bars in condition (2), and varied across bars from $L_0 - m$ to $L_0 + m$ in condition (1&2). Since the 15-bar display contains 1.25 cycle of the stationary grating, the space-averaged luminance of this grating is not necessarily zero and depends on which 1.25-cycle segment is displayed. We randomly displayed two segments on different trials that differed by π. Thus, across trials, effects of nonzero space-averaged luminance of the stationary grating would cancel each other. Second, we used 15 bars that had one third the usual overall width, resulting in a display that had the same width as usual (0.22 deg) but was spatially a finer approximation to a sine wave. Third, we used two temporal frequencies (4.2 and 16.7 Hz). Fourth, we made from 576 to 1728 observations in each of the six [(1) versus (1&2) versus (2), and ω] conditions.

2. Results

Data were similar to those in Experiment 5 (Fig. 14), although subject NB showed a slight but insignificant trend for performance in condition (1&2) to be worse than in conditions (1) and (2). Overall, the data of Experiments 5 and 6 show that adding a stationary sine-wave background (of the same or different spatial frequency than the moving grating) has no effect on movement detection other than the small effect that is to be expected from the addition of (uniform) background luminance. These data further confirm the original and elaborated Reichardt models. Experiments 5 and 6 each lead to rejection of the following models: (1) spatial correlation analysis, (2) local brightness matching, (3) Marr–Ull-

man zero crossings, (4) spatiotemporal gradient matching; Experiment 6 requires rejection of spatial phase analysis.

G. Experiment 7. Adding Homogeneous Flicker

The previous experiment showed that adding a stationary grating that has the same spatial frequency as a moving grating has little, if any, effect on the accuracy of direction judgments. We had predicted this finding on the basis of two properties that are shared by Reichardt's original model and the elaborated model, namely, (1) segregation of temporal frequencies and (2) having a zero response to a stationary pattern. In the present experiment, we investigate the effects of a dual manipulation: the addition of homogeneous flicker that has the same temporal frequency as a moving sine wave.

1. Predictions of the Original Reichardt Model

The Reichardt model predicts that addition of homogeneous flicker to a moving sine wave should affect motion perception. Even though, as can be easily shown, the response to flicker presented alone is zero, the response to the sum of two luminance modulations that have the same temporal frequency is not equal to the sum of the responses to the luminance modulations. Obviously, merely predicting that adding homogeneous flicker affects performance would not constitute a critical test of the elaborated Reichardt model. However, it turns out that, under appropriate conditions, the Reichardt model predicts that addition of homogeneous flicker will reverse the direction of perceived motion.

Figure 15 illustrates the logic underlying the prediction of flicker-produced motion reversal in the Reichardt model. Figure 15A contains an ordinary moving-grating display L with $\varphi = \pi/6$ (the control display). That is, $L_j(t) = L_0 + m\sin(2\pi\omega t - \vartheta_j)$, with $\vartheta_j = -0.33\pi, -0.17\pi, 0, 0.17\pi$, and 0.33π, for $j = 1, \ldots, 5$. Figure 15B shows homogeneous flicker L' that has the same temporal frequency ω but with twice the amplitude ($2m$) and that has a relative phase (φ_{rel}) of π (Fig. 15B). The relative phase of flicker φ_{rel} is defined to be the phase of L' relative to the phase of the center (third) bar of L. Thus $L_j'(t) = L_0 + 2m\sin(2\pi\omega t - \varphi_{rel})$ for $j = 1, \ldots, 5$; $\varphi_{rel} = \pi$. Then the resulting reversal display, $L'' = L + L' - L_0$ in Fig. 15C, is obtained by adding the time-varying components of L and L' while keeping the average luminance at L_0. Thus one could say that L'' consists of a control component (the time-varying part of L) and a flicker component (the time-varying part of L'). The same logic, with the roles of space and time interchanged, was used in the previous experiment, in which we tested the spatial phase analysis model (Fig. 13). In the context of the present experiment, we spell out the mathematics in full detail. As can be seen in Fig. 15C, L'' consists of temporal sinusoids that again have the same temporal frequency ω but an ascending temporal phase line and hence phase differences that are opposite in sign to those in Fig. 15A; also note that the amplitudes vary across the display. Specifically, $L_j'' = L_0 + m_j'' \sin(2\pi\omega t - \vartheta_j'')$, where $\vartheta_j'' = 1.17\pi, 1.13\pi, \pi, 0.87\pi$, and 0.83π, and $m_j'' = 1.73, 1.24, 1.00, 1.24$, and $1.73m$ for $j = 1, \ldots, 5$.

The compound display of Fig. 15C illustrates the following property of sinusoids:

$$\sin(\omega t + x) + K \sin(\omega t + \pi) = A(x)\sin[\omega t + B(x)]. \quad (23)$$

Here $A(x) \geq 0$, and $B(x) = \tan^{-1}\{\sin(x)/[\cos(x) - K] + \pi$. What is critical is that $B(x)$ is a nonmonotonic function of x. In fact, it follows that, by restricting x to an interval in which $B(x)$ is decreasing in x, we can reverse the sign of the phase differences between adjacent bars. The back-and-forth path of the spatial phase path in Experiment 6 corresponds to the nonmonotonicity of $B(x)$ in the present experiment. In Experiment 6, however, we did not restrict t to an interval in which the spatial phase path is descending, because this would
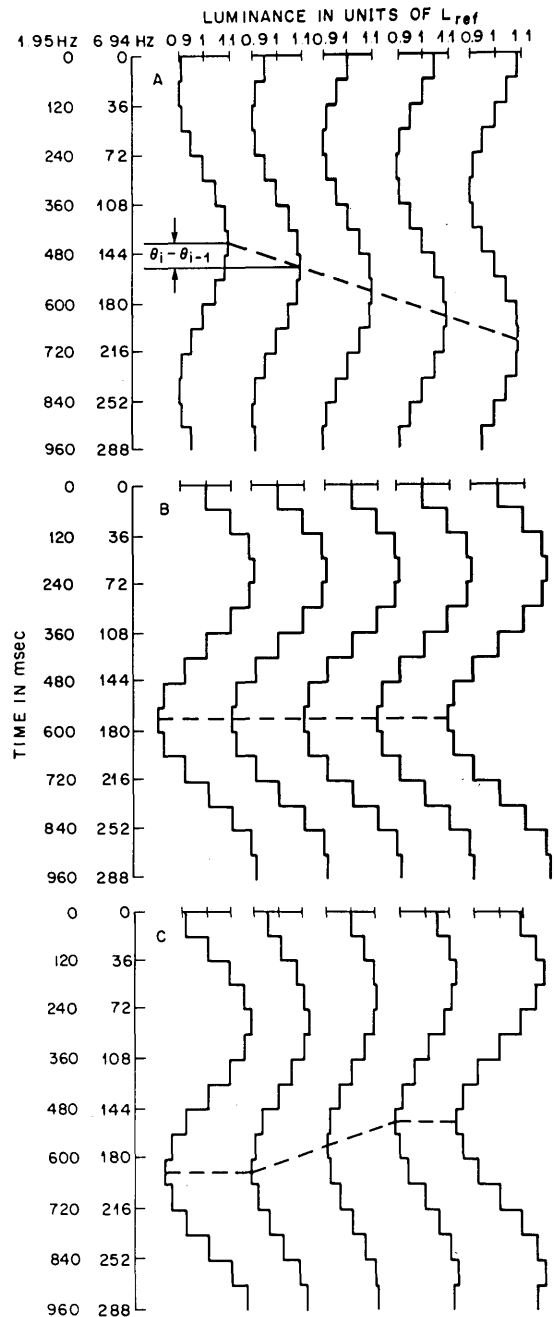

Fig. 15. Displays used in Experiment 7. A, sinusoidal modulation function with a synchrony of $\pi/6$ rad. B, sinusoidal modulation function with asynchrony of 0 rad (homogeneous flicker) and same temporal frequency ω but twice the modulation m used in A. C, result of adding flicker in B to display in A. Note that asynchronies are opposite in sign to those in A, whereas modulation m_j varies across fields.

require display durations of one third of a temporal cycle; this would violate the assumption of temporal periodicity on which all our analyses are based.

According to the original Reichardt model, the sign of detector output depends on the sine of the temporal phase difference in the input channels [Eq. (8)], $\sin(\gamma_{\text{right},n} - \gamma_{\text{left},n})$. In the control display, these differences begin at $+0.17\pi$ (for adjacent bars) and increase to $+0.67\pi$ (for bars 1 and 5). In the reversal display of Fig. 15C, the differences begin at -0.034π (for bars 1 and 2 and for bars 4 and 5) and decrease to -0.33π (for bars 1 and 5). No matter from which pair of bars a detector derives its input, the temporal phase difference is between 0 and π in the control display and between 0 and $-\pi$ in the reversal display. Hence the control and the reversal displays yield detector outputs that are opposite in sign. This is the case for each individual detector. Hence our linking assumption implies that direction of perceived motion should reverse. Note, however, that the control and the reversal display differ in more respects than just the sign of the phase differences. First, the phase differences have different magnitudes. Second, the amplitudes m and m_j'' are unequal. Hence, in contrast to Experiment 2, the original Reichardt model does not predict that reversal should be perfectly symmetrical, i.e., that performance in the reversal condition should fall below the 50% mark by precisely the same amount by which performance in the control condition should exceed 50%.

The original model makes additional predictions. First, when added homogeneous flicker is in phase with the central bar in Fig. 15A ($\varphi_{\text{rel}} = 0$), phase differences are not reversed. Thus superimposition of this type of flicker should not lead to reversal of motion direction. Second, when added homogeneous flicker has a different temporal frequency than the moving grating [$\omega' \neq \omega$, Eq. (23)], no reversal of motion direction is predicted at any relative phase. This follows from the property of segregation of temporal frequencies of detector outputs and from the fact that detector output to homogeneous flicker is zero.

2. Predictions of the Elaborated Reichardt Model

The elaborated Reichardt model makes the same predictions for added, homogeneous flicker as the original model. The prediction is based on an analysis of the primary spatiotemporal sine-wave components of the displays and can be summarized as follows: The control display L (Fig. 15A) is dominated by $\sin(2\pi\omega t + 2\pi 1.89x)$. The homogeneous flicker L' (Fig. 15B) is simply $\sin(2\pi\omega t - \pi)$. The reversal display L'' (Fig. 15C) is dominated by $\sin(2\pi\omega t - 2\pi 1.19x)$. In other words, the displays in Figs. 12A and 12C are dominated by sine waves that move in opposite directions. Hence the elaborated model, like the original Reichardt model, predicts that opposite directions of motion are perceived for the control and reversal displays. Second, when homogeneous flicker is added in phase with the central bar of the moving grating (Fig. 15A), the dominant sine wave of the compound stimulus is $\sin(2\pi\omega t + 2\pi 1.19x)$. Hence no reversal should be observed. Third, when the flicker has a different temporal frequency from the moving sine wave, no effects should be observed, because the elaborated Reichard model, like the original Reichardt model, has the property of segregation of temporal frequencies and has a zero response to homogeneous flicker.

Reversal of perceived direction because of addition of flicker casts considerable doubt on several of the alternative models. A homogeneous change in the luminance of a frame has no effect on the spatial correlation with the preceding frame, nor does this change affect the spatial phase of the most prominent spatial Fourier component. [Spatial Fourier integral analysis of either the entire display (i.e., the five bars and the dark surround) or only of the five bars shows that motion paths of spatial Fourier components are insignificantly affected and certainly not reversed in direction by homogeneous flicker.]

3. Method

The control display, a moving grating, is illustrated in Fig. 15A and was discussed above. The value of $\pi/6$ for between-bar temporal phase difference φ yields a spatial frequency of 1.89 cpd. The temporal frequency (ω') of added flicker was $\omega/2$, ω, and 2ω. The following values of φ_{rel} were tested: for $\omega' = \omega/2$, $\varphi_{\text{rel}} = 0$ and π; for $\omega' = \omega$, $\varphi_{\text{rel}} = 0$ and π; and for $\omega' = 2\omega$, $\varphi_{\text{rel}} = 0$ and $\pi/2$ (note that here $\varphi_{\text{rel}} = 0$ is equivalent to $\varphi_{\text{rel}} = \pi$).

The seven conditions (the control display and the six added-flicker displays) were presented at two temporal frequencies (2.08 and 6.94 Hz), with a modulation of 0.128. Two observers made 96 observations in each of the 12 added-flicker conditions and 576 observations in each of the two control conditions.

4. Results

There was an overall effect of temporal frequency in the control condition (dashed lines in Fig. 16) consistent with the temporal MTF's obtained in Experiment 4 (Fig. 11). In the reversal condition ($\omega' = \omega$, and $\varphi_{\text{rel}} = \pi$), the apparent direction of motion was reversed in all cases, although the reversal was not quite symmetrical. That is, the reversed stimulus was opposite (as expected) but not quite equal in effectiveness to the control. Similarly, the in-phase flicker ($\varphi_{\text{rel}} = 0$) enhanced the perception of direction in the control direction (in three of four cases), but the magnitude of the enhancement is difficult to estimate because of ceiling effects. Added flicker at frequencies different from the moving grating has little effect in three of four cases (as predicted). The exceptional case (subject NB, 6.94 Hz) can be best understood by assuming that the subject's performance was disturbed in some general way by added flicker in all the conditions. Once this overall
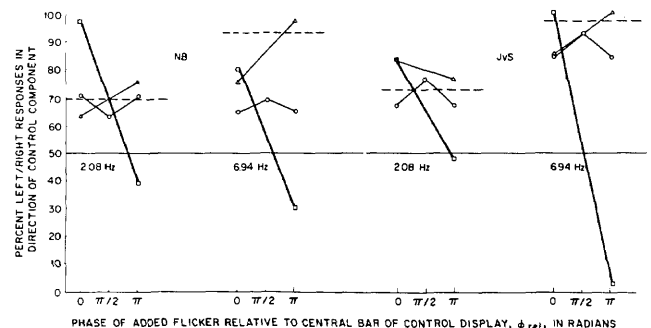

Fig. 16. Results from Experiment 7. Effects of adding homogeneous flicker. Dotted horizontal lines indicate performance without added flicker. The remaining lines indicate performance with added flicker. The flicker frequency is $2\omega(\bigcirc)$, $\omega(\square)$, or $\omega/2(\triangle)$, where ω is the temporal frequency of the display without added flicker.

deficit has been parceled out, these data became consistent with the rest. The magnitude of the reversal effect is greater than the magnitude of the enhancement effect (distance of endpoints of heavy lines from dashed control line, Fig. 16). The reversed stimuli are less strongly perceived (in the opposite direction) than the control gratings for many reasons. The differences in between-bar phase difference (ϑ_j versus ϑ_j'') and in modulation amplitude (m versus m'') have already been pointed out. Additionally, for the specific conditions tested, by extrapolating from Fig. 11, we can infer that, at the temporal frequencies used, sensitivity is lower for the spatial frequency in the reversal condition (1.19 cpd) than for the spatial frequency in the control condition (1.89 cpd). In addition, Fig. 11 shows that reduced sensitivity to low spatial frequencies should be more pronounced at lower temporal frequencies, as we, in fact, found in the present experiment. Second, the fraction of a spatial cycle contained in an individual frame was smaller in the reversal than in the control condition. It is easy to show that the detector output increases with the size of this fraction.

In summary, the $\omega/2$ and 2ω data strongly support the predictions based on the property of segregation of temporal frequencies that are shared by the original and the elaborated Reichardt models. The prediction of reversal of perceived direction by addition of homogeneous flicker to a moving grating, a prediction based on nonlinearity of the Reichardt model, is also strongly supported.

4. DISCUSSION

We have proposed a model of direction-sensitive units in human vision that is a modified and elaborated version of Reichardt's model. The model analyzes a small patch of the visual field, and its response is additive when input components have different temporal frequencies.

The model was tested in experiments in which the subjects viewed adjacent vertical bars with independently (typically sinusoidally) modulated luminances. The experiments confirmed several fundamental properties of the model. Experiment 1 demonstrated that motion detection involves sine-wave analysis in the temporal domain. Experiments 2 and 3 showed that, when the between-bar phase difference is $\pi/2$, detection performance depends on the product of adjacent bar amplitudes. Experiment 4 rejected the point-input assumption of the original Reichardt model, provided evidence that input receptive fields have both on and off areas, and demonstrated that the detector population is heterogeneous. Detectors sensitive to high temporal frequencies are less sensitive to high spatial frequencies and vice versa. Experiments 5 and 6 showed that addition of a stationary pattern has little or no effect on performance, even when the amplitude of the pattern exceeds that of the moving pattern and when the spatial frequencies of the stationary and moving pattern are the same. Experiment 7 showed that the direction of perceived motion can be reversed by adding homogeneous flicker, but only when its temporal frequency is the same as that of the moving sine wave and when it has the appropriate temporal phase relationship with the moving sine wave.

These results cast doubt on a number of alternative models. In particular, the final three experiments demonstrate that motion perception does not involve the frame-to-frame

comparison processes that are hypothesized by several of these models. The reason is that addition of stationary patterns changes pattern information within individual frames but has no effect on performance, whereas addition of homogeneous flicker does not affect directional information in frames but is shown to reverse the perceived direction of motion.

Our model differs fundamentally from models that involve frame-to-frame comparison, because, rather than comparing two spatial luminance patterns at two instants in time, it compares two temporal luminance patterns at two locations.

A. Time Averaging and Adaptation

1. Infinite Time Averaging
Perhaps the most implausible assumption of the Reichardt model is infinite time averaging. The role of time averaging in the model is important, because it removes all time-varying components of the subunits' outputs. However, because time averaging requires integration over an infinite amount of time, this operation cannot be implemented in real-world systems. If, as suggested by Foster,[28] TA is replaced by low-pass temporal filtering, these subunit outputs will fluctuate around their average values. If the time-varying components in the input to these low-pass temporal filters are sufficiently fast, or if viewing time is sufficiently long to allow the observer to view several cycles of each component and perform a type of mental time averaging, replacement of the TA units by low-pass filters will have few observable effects. However, if there is a slow, time-varying component and only one segment of its temporal cycle is shown, then significant departures from the model should be observed. We are currently working on this issue.

2. Linearity
The assumption of linearity of TF, TA, and of any other temporal filters that could be included in the model is a weak spot. Because we have used small modulation depths, we have not taxed this assumption. It would be of interest to see how large the modulation depth m has to be for linearity to break down. A related issue is the dependency of detection performance on average luminance level L_0 demonstrated in Experiments 5 and 6. This finding clearly requires a further elaboration of the elaborated Reichardt model to incorporate stimulus quantum noise (the amplitude of which is related to stimulus level) as well as gain-control adaptation in the input channels.

B. Physiological Considerations
Cats have cortical neurons that exhibit several of the fundamental properties of the elaborated Reichardt model. Movshon et al.[36] found complex cells in area 18 that have the following properties. First, they are directionally selective. Second, these cells derive their inputs from linear subunits. Third, the receptive fields of the subunits consist of several on and off areas. Fourth, the complex cells have essentially zero sensitivity to stationary patterns (their temporal-frequency sensitivity is band-pass). Fifth, there are facilitory (and, to a lesser extent, inhibitory) interactions between the subunits. Sixth, these interactions cause the complex cell to respond in certain ways that are characteristic of the elabo-

rated Reichardt model: the response to a moving sine wave consists primarily of a steady-state component, whereas the response to a counterphase grating consists primarily of the second harmonic. This is precisely what the modified model predicts, provided that (a) recordings are made before time averaging (or time averaging is imperfect; see above), (b) the receptive-field arrangement is given by Eq. (12), and (c) the temporal phase shift $\delta_\omega = \pi/2$. Finally, there is evidence that most cells in the striate cortex conform to the separability hypothesis, according to which the optimal temporal frequency is independent of the spatial frequency of a moving sine wave, and vice versa.[11]

These considerations indicate that the elaborated Reichardt model is not inconsistent with current physiological facts. However, it is not necessary for physiological mechanisms to have precisely the same structure as the elaborated Reichardt model (e.g., Fig. 2). As we shall see in the next sections, there exist several versions of the elaborated Reichardt model (with different components) that are functionally equivalent in that they perform the same net computation.

C. Equivalent Forms

1. Shunting Inhibition versus Multiplication
It is easy to show[37] that the behavior of a detector's subunits, and hence that of the detector as a whole, is not affected if we simultaneously (a) add a high-pass temporal filter somewhere before the multiplication unit and (b) replace this multiplication unit by shunting inhibition (i.e., multiplication of the high-pass-filtered input with one minus the input received from TF). The latter change has the effect of reversing the preferred direction. The underlying logic is that inhibition in the nonpreferred direction is equivalent to facilitation in the preferred direction. The high-pass input filter is needed to eliminate steady-state components in the output of the inhibition unit.

2. Direct Subunit Access and Generalized Subtraction
In the elaborated Reichardt model, the subject's response is based not on the outputs from the individual subunits but on the difference between these outputs. However, one can imagine a different scheme in which the responses are based directly and completely on subunit outputs rather than on their pairwise differences. In fact, one can imagine a scheme in which subunits tuned to rightward and leftward motion have no direct, "physical" connection with each other and occur in pairs only in the formal sense that for every leftward-tuned detector we can find elsewhere in the system a rightward-tuned detector that is its complete mirror image. In this new scheme, it still would be reasonable to assume that the probability of giving a right response would increase in the output from the right subunit and decrease in the output from the left subunit but would not necessarily be equal to the difference between these two outputs. Formally, in the new scheme for any stimulus $L(x, t)$, $P[\text{right}|L] = V[g(y_{\text{right},3,1}, y_{\text{left},3,1}), \ldots, g(y_{\text{right},3,M}, y_{\text{left},3,M})]$. Here, $y_{H,3,i}$ denotes the output of the ith subunit tuned to direction H (left, right); g is a generalized subtraction operation that has no restrictions other than that it increases in the first argument, decreases in the second argument, and has the (antisymmetry) property that $g(a, b) = -g(b, a)$; and V is a reasonable voting rule, i.e., nondecreasing in all of its M arguments.

The question now arises under what conditions this generalized subtraction scheme is equivalent to the old simple-difference scheme. That is, what, if any, amendments in the subunits do we have to make, and which properties must a display L have so that for any g having the above properties, the output of detector D_i (according to the simple-difference scheme), $y_{4,i}$, is monotonically related (by some monotonically increasing function H) to $g(y_{\text{right},3,i}, y_{\text{left},3,i})$? Here, $y_{4,i}$ is defined as $y_{\text{right},3,i} - y_{\text{left},3,i}$, as always.

Whether generalized subtraction is equivalent to simple difference depends on the following conditions.

(1) The scheme must incorporate the high-pass temporal filter, as indicated above.

(2) Temporal phase shift $\delta_{n\omega}$ [Eq. (5)] must be $\pi/2$ for all n.

(3) The input receptive fields must be of the form described in Eq. (12), or, to be more exact, $r_{\text{left}}(x)$ must be equal to the convolution of $r_{\text{right}}(x)$ with an arbitrary function $a(x)$ that has the property of antisymmetry [i.e., $a(-x) = -a(x)$].

(4) For each temporal frequency ω, L contains at most one spatiotemporal Fourier component with that frequency.

It can be shown that $y_{\text{right},3,i} = -y_{\text{left},3,i}$, i.e., the outputs from the left and right subunits are completely equivalent when either conditions (1), (2), and (3) are satisfied or conditions (1) and (4) are satisfied. If we now let $H(x) = g(x/2, -x/2)$, it follows that H is monotonically increasing, and $g(y_{\text{right},3,i}, y_{\text{left},3,i}) = H(y_{4,i})$; this shows that the generalized subtraction g is monotonically related to $y_{4,i}$.

Note that the generalized subtraction scheme includes the important special case in which negative outputs from a subunit are truncated at zero. [If we let $g(a, b) = a$ if $a > -b$ and $g(a, b) = b$ if $a < -b$, it can be easily demonstrated that, thus defined, g has the desired properties.]

The existence of several equivalent forms and the wide range of conditions for equivalence show that the behavior of the model does not depend on details, such as the precise type of comparison operation (multiplication versus shunting inhibition), the voting rule, and the nature of the receptive fields and temporal filters. Most of the model's behavior stems from its deep structure, i.e., the nonlinear comparison of temporal luminance modulations in adjacent locations.

D. Detection of Counterphase Gratings
In our task, subjects were to determine direction of motion. A more common task, however, consists of detecting the presence of a (moving or nonmoving) spatial pattern. As stated earlier, nonmoving patterns, such as stationary patterns, homogeneous flicker, and counterphase gratings, lead to zero detector output. Hence these patterns cannot be detected on the basis of our hypothetical detectors. This leaves us with two possibilities. First, detection of nonmoving patterns involves a population of units that are not direction selective. Alternatively, detection of nonmoving patterns involves direct access to subunit output before subtraction. For example, even though the detector's response to counterphase gratings is zero, the subunit's responses are in general not zero. [Note that counterphase gratings do not have condition (4) needed for generalized subtraction to be equivalent with the standard subtraction scheme.]

This brings us to a current theoretical issue, namely, the explanation of the fact that contrast sensitivity is lower for a counterphase grating than for a moving sine wave with the same temporal and spatial frequency. Levinson and Sekuler[7] and Watson et al.[8] discuss the elegant hypothesis that detection of a counterphase grating is mediated by channels that are tuned to its moving components, i.e., to $S(m/2, d, f, \omega)$ and $S(m/2, -d, f, \omega)$, where S is a grating stimulus with spatial frequency f and temporal frequency ω, d is the direction of movement, and m is the modulation depth of the counterphase grating. It follows that, for a single channel to reach threshold, a moving sine wave needs only one half of the modulation depth that a counterphase grating needs. Watson et al.[8] argue that, under reasonable linking assumptions, this hypothesis explains why contrast sensitivity is lower for counterphase gratings than for moving sine waves.

Although the subunits in the Reichardt model could be interpreted as channels that are tuned to leftward and rightward motion, the model suggests quite a different explanation. It does not decompose a counterphase grating into its moving sine-wave components. Stimulus components that have the same temporal frequency, as is the case for the moving sine-wave components of a counterphase grating, interact in complicated ways in our detector system. The response $y_{3,H}$ of subunit H to a counterphase grating is generally not equal to the response to the moving sine-wave component to which the subunit is tuned. The relative magnitude of the subunit's responses to moving sine waves and counterphase gratings depends on the properties of SF and TF. For example, when we have the receptive-field arrangement expressed in Eq. (12), the ratio of the sensitivities of a subunit to moving sine waves and counterphase gratings depends on $\tan \delta_\omega$, where δ_ω is the temporal phase shift. When TF is a first-order filter with time constant τ, the ratio is $2\omega\tau$. This ratio can be equal to 2 for at most one temporal frequency ω.

In summary, our subunits cannot play the role needed for the explanation given by Levinson and Sekuler and by Watson et al. According to the Reichardt model, a counterphase grating is not decomposed into its component sine waves. Any regularity in the relation between thresholds for counterphase gratings and for moving gratings either does not reflect the workings of Reichardt-type detectors (and must be a result of other factors) or it must be a result of accidental properties of SF and TF.

E. Spatiotemporal Fourier Analysis

In the introduction, we stated the general notion that the human visual system performs some type of spatiotemporal Fourier analysis. In this section, we spell out in more detail to what extent and how the elaborated Reichardt model performs this spatiotemporal Fourier analysis.

In the context of direction-discrimination tasks, we mean by spatiotemporal Fourier analysis the following two-step process. First, the display is decomposed into Fourier components (spatiotemporal sine waves). Second, a voting rule is applied to the components to decide which direction has received the most weight, i.e., which direction has the largest Fourier component amplitudes. We use V' to distinguish it from the voting rule V, which is applied to detector outputs. It seems reasonable to require that V' should be increasing in each argument. Thus the sign of V' indicates direction, whereas its magnitude indicates the strength of the vote.

First, consider displays in which all Fourier components have different temporal frequencies. For a single, elaborated Reichardt detector, the response is the weighted sum of squared amplitudes; the sign of each term in this sum corresponds to the direction of the Fourier component. This follows from segregation of temporal frequencies [Eqs. (8a) and (8b)] and from the absence of temporal and spatial aliasing (Sections 1.C.1 and 1.C.2). Thus, for these displays, a single, elaborated Reichardt detector performs the spatiotemporal Fourier analysis that we described above with an additive voting rule V'.

For an entire system of detectors, a voting rule V is applied to combine the individual detector responses. Because V is strictly increasing in each argument (detector output) and each detector output is increasing in each Fourier component amplitude, V is strictly increasing in each Fourier component amplitude. Thus, for displays in which all Fourier components have different temporal frequencies, the elaborated Reichardt model indeed performs a spatiotemporal Fourier analysis.

Second, consider displays in which some Fourier components have the same temporal frequency. Now the property of segregation of temporal frequencies does not apply, and the elaborated Reichardt model does not, in general, perform a Fourier analysis. In particular, as predicted by the elaborated Reichardt model, Experiment 7 (and, in particular, an informal replication with a fixation point that provides an attentional rather than a physical window) showed that perceived direction is location dependent and thus does not depend only on the amplitudes of the Fourier components. The only condition under which the elaborated Reichardt model performs a Fourier analysis of shared temporal-frequency displays is when Fourier components with the same temporal frequencies have sufficiently different spatial frequencies to stimulate nonoverlapping sets of detectors.

We now argue that for the remaining displays, i.e., displays that have components with shared temporal frequencies and similar spatial frequencies, our model's behavior may be described as that of performing a local spatiotemporal Fourier analysis. By this we mean that for a given location we first apply some attentional window to the display and then apply spatiotemporal Fourier analysis. What is the effect of a window on Fourier components that share the same temporal frequency and that have similar spatial frequencies? The usual effect is to spread out their spatial-frequency spectra in such a way that the combined spatial-frequency spectrum has only one dominant peak. (This happened in Experiment 7, in which the spatial-frequency spectra of a physically windowed drifting grating and uniform flicker combined to form a spectrum with one prominent peak for a direction of movement opposite that of the drifting grating.) In effect, a window causes Fourier components with the same temporal frequency and with similar spatial frequencies to become unresolvable. Then, however, we have a display in which there are no shared temporal-frequency components, so that our model performs a spatiotemporal Fourier analysis.

Analysis of a linear, local, spatiotemporal Fourier analysis model, recently proposed by Watson and Ahumada,[38] provides further evidence of the profound relation between the elaborated Reichardt model and the notion of local, spatiotemporal Fourier analysis. Their model is incomplete in that it does not specify how the ultimate left–right response

is based on the time-varying detector output. We embellished their model with an appropriate linking assumption: squaring and time averaging of the detector outputs and subsequently subtracting outputs of left- from right-tuned detectors. The resulting model can be shown to be fully equivalent to a special case of elaborated Reichardt model.

To summarize, for all displays except for those that have components with shared temporal frequencies and similar spatial frequencies, the elaborated Reichardt model performs a spatiotemporal Fourier analysis; for the remaining displays, the elaborated Reichardt model's occasional failure to perform a spatiotemporal Fourier analysis is resolved by performing a local Fourier analysis. As the empirical evidence accumulates, the range of variation in models that describe short-range motion perception is becoming restricted to models that do not differ essentially from the elaborated Reichardt model.

ACKNOWLEDGMENTS

REFERENCES

1. O. Braddick, "A short-range process in apparent motion," Vis. Res. **14**, 519–529 (1974); G. Westheimer, "The spatial sense of the eye," Invest. Ophthalmol. Vis. Sci. **18**, 893–912 (1979).
2. S. M. Anstis, "Apparent movement," in *Handbook of Sensory Physiology, Vol. VIII: Perception*, R. Held, H. W. Leibowitz, and H.-L. Teuber, eds. (Springer-Verlag, New York, 1977).
3. A. J. Pantle and L. Picciano, "A multi-stable movement display: evidence for two separate motion systems in humans," Science **193**, 500–502 (1976).
4. O. J. Braddick, "Low-level and high-level processes in apparent motion," Philos. Trans. R. Soc. London Sect. B **290**, 137–151 (1980).
5. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Physiol. (London) **197**, 551–566 (1968).
6. H. R. Wilson and J. R. Bergen, "A four mechanism model for threshold spatial vision," Vision Res. **19**, 19–33 (1979).
7. E. Levinson and R. Sekuler, "The independence of channels in human vision selective for direction of movement," J. Physiol. (London) **250**, 347–366 (1975).
8. A. B. Watson, P. G. Thompson, B. J. Murphy, and J. Nachmias, "Summation and discrimination of gratings moving in opposite directions," Vision Res. **20**, 341–347 (1980).
9. E. Adelson, "Some new motion illusions, and some old ones, analyzed in terms of their Fourier components," presented at the meetings of the Association for Research in Vision and Ophthalmology, Sarasota, Florida, 1982.
10. W. Reichardt, "Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems," Z. Naturforsch. Teil B **12**, 447–457 (1957); W. Reichardt and D. Varju, "Uebertragungseigenschaften im Auswertesystem fuer das Bewegungssehen," Z. Naturforsch. Teil B **14**, 674–689 (1959); W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, W. A. Rosenblith, ed. (Wiley, New York, 1961).
11. Equation (10) implies the separability hypothesis [H. R. Wilson,

"Spatiotemporal characterization of a transient mechanism in the human vision system," Vision Res. **20**, 443–452 (1980)] according to which the spatiotemporal frequency response of a detector is the product of the spatial- and the temporal-frequency responses. However, Eqs. (11)–(22) can still be derived if we replace Eq. (10) by a weaker form: $y_{H,0} = \Sigma_{n=0}^{\infty} \int r_{n\omega}(x) L_n(x, t) \, dx$, which does not imply separability.

12. D. Gabor, "Theory of communication," J. Inst. Electr. Eng. **93**, 429–457 (1946).
13. S. Marcelja, "Mathematical description of the responses of simple cortical cells," J. Opt. Soc. Am. **70**, 1297–1300 (1980).
14. For example, J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," Vision Res. **20**, 837–846 (1980); C. R. Carlson, R. W. Klopfenstein, and C. H. Anderson, "Spatially inhomogeneous scaled transforms for vision and pattern recognition," Opt. Lett. **6**, 386–388 (1981).
15. D. A. Pollen and S. F. Ronner, "Phase relationships between adjacent simple cells in the visual cortex," Science **212**, 1409–1411 (1981).
16. In fact, Eq. (15) holds for any pair of symmetric receptive fields, including ones that are completely low-pass. This implies that the detector can have band-pass characteristics even when its input receptive fields are low-pass and that, thus, a psychophysical band-pass response, to the extent that it is based on Reichardt-type detectors, does not imply the existence of single cells with band-pass linear receptive fields.
17. The critical property needed for the specific voting rule V is that, when detector output z_i is of the form $z_i = x \xi_i$ (where ξ_i is a factor specific to detector D_i), then V is monotonic in x. We need this property to test the multiplicative law [Eq. (19)], where the role of x is played by the detector-independent term $m_{odd} m_{even}$ and the role of ξ_i is played by the detector-dependent term $\Sigma_{k=1,3,5,\ldots}^{F-1} \Sigma_{j=1}^{F-k} p(\vartheta_{j+k} - \vartheta_j) A_{jk}$. The voting rule used to generate predictions is

$$V(z_1, \ldots, z_M) = q \left[\sum_{i=1}^{M} k \, \text{sign}(z_i) |z_i|^p \right],$$

where q is a function that ranges between 0 and 1, is antisymmetric around 0.50 [i.e., $q(z) - 0.50 = 0.50 - q(-z)$], and is strictly increasing; k and p are arbitrary positive constants. This rule includes both the additive case (for $p = k = 1$), in which the response depends on the sum of the detector outputs, and the maximum case [for $p \to \infty$, $k = 1$, and $q(z) = z^{1/p}$], in which the response depends on the maximum of the detector outputs (as in threshold models).
18. J. A. Leese, C. S. Novak, and V. R. Taylor, "The determination of cloud pattern motions from geosynchronous satellite image data," Pattern Recognition **2**, 279–292 (1970).
19. E. A. Smith and D. R. Phillips, "Automated cloud tracking using precisely aligned digital ATS pictures," IEEE Trans. Comput. **C-21**, 715–729 (1972).
20. R. C. Lo and J. A. Parikh, "A study of the application of Fourier transforms to cloud movement estimation from satellite photographs," Comput. Sci. Tech. Rep. TR-242 (University of Maryland, College Park, Md., 1973).
21. S. M. Anstis and B. J. Rogers, "Illusory reversal of visual depth and movement during changes of contrast," Vision Res. **15**, 957–961 (1975).
22. S. Ullman, "Analysis of vision motion by biological and computer systems," Computer **14**, 57–69 (1981).
23. J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images in television signals," Comput. Graphics **4**, 311–327 (1975).
24. C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," Comput. Graphics **9**, 301–315 (1979).
25. D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing," Proc. R. Soc. London Sect. B **211**, 151–180 (1981).
26. D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. London Sect. B **207**, 187–217 (1980).
27. J. F. Schouten, "Subjective stroboscopy and a model of visual movement detection," in *Proceedings of the Symposium on Models of the Perception of Speech and Visual Form* (MIT, Cambridge, Mass., 1967).

28. D. H. Foster, "A model of the human visual system in its response to certain classes of moving stimuli," Kybernetik 8, 69–84 (1971).
29. G. Sperling, "Movement perception in computer-driven visual displays," Behav. Res. Methods Instrum. 8, 144–151 (1976).
30. A. Korte, "Kinematoscopische Untersuchungen," Z. Psychol. 72, 193–296 (1915).
31. D. H. Kelly, "Motion and vision. II. Stabilized spatiotemporal threshold surface," J. Opt. Soc. Am. 69, 1340–1349 (1979).
32. D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," Vision Res. 22, 479–484 (1982).
33. These simulations consisted of calculating the value of $\sum_{j=1}^{3} A_{j2}$, with x_c (location of detector center relative to display center), $x_{\text{right}}-x_{\text{left}}$ (receptive-field center distance), and scale parameter σ factorially taking values between 0.05 and 105.2 by steps of a factor of 2. We did this for $r_H(x) = \exp[-(x - x_H)^2/\sigma^2]$, $\exp(-|x$

$- x_H|/\sigma)$ uniform with width parameter σ, triangular, and semicircular.
34. R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, *Statistical Inference under Order Restrictions; The Theory and Application of Isotonic Regression* (Wiley, New York, 1972).
35. B. J. Murphy, "Pattern thresholds for moving and stationary gratings," Vision Res. 18, 521–530 (1978).
36. J. A. Movshon, I. D. Thompson and D. J. Tolhurst, "Receptive field organization of complex cells in the cat's striate cortex," J. Physiol. 283, 79–99 (1978).
37. J. Thorson, "Small-signal analysis of a visual reflex in the locust. II. Frequency dependence," Kybernetik 3, 53–66 (1966).
38. A. B. Watson and A. J. Ahumada, "A Look at Motion in the Frequency Domain," NASA Tech. Mem. 84352 (National Technical Information Service, Springfield, Va., 1983).