Perception, 1996, volume 25, pages 1263–1280

Transient–sustained input to directionally selective motion mechanisms

Walter F Bischof
Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada
Adriane E Seiffert
Department of Psychology, Harvard University, Cambridge, MA 02138, USA
Vincent Di Lollo¶
Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;
e-mail: enzo@cortex.psych.ubc.ca
Received 12 December 1995, in revised form 28 April 1996

Abstract. The characteristics of the sustained input to directionally selective motion sensors were examined in three human psychophysical studies on directional-motion discrimination. Apparent motion was produced by displaying a group of dots in two frames (F1 and F2), where F2 was a translated version of F1. All stimuli included parts that contained both F1 and F2 (combined images) and parts containing only F1 or F2 (single images). All displays began with a single image (F1), continued with the combined image, and ended with F2. Six durations of single and of combined images (10, 20, 40, 80, 160, or 320 ms) were crossed factorially. As the duration of the single image was increased, perception of directional motion first improved, and then declined at longer durations. This outcome contrasted with the monotonic increment obtained in earlier studies under low-luminance conditions. To account for the entire pattern of results, earlier models of the Reichardt motion sensor were modified so as to include a mixed transient–sustained input to one of the filters of the sensor. Predictions from the new model were tested and confirmed in two experiments carried out under both low-luminance and high-luminance viewing conditions.

1 Introduction

Apparent motion of a group of random dots can be produced in a display sequence comprising a leading frame (F1) and a trailing frame (F2), with F2 being the same as F1 except for a small unidirectional displacement of each dot. Under appropriate spatiotemporal conditions, motion is seen in the direction of the displacement. Theoretical models have employed hypothetical low-level sensors as sources of motion signals. Perhaps the best known are the models proposed by Reichardt (1961), Adelson and Bergen (1985), van Santen and Sperling (1985), and Watson and Ahumada (1985).

In this class of models, the production of a motion signal depends on an interaction between two transient responses, one triggered by the onset of F1, the other by the onset of F2. Given that transient responses are relatively brief, it follows that the onsets of F1 and F2 must occur closely in time if an interaction between them is to take place. Contrary to this expectation, recent neurophysiological and psychophysical evidence indicates that motion signals can be produced even if the onsets of F1 and F2 are separated by temporal intervals that exceed the duration of transient responses (Baker and Cynader 1994; Bischof and Di Lollo 1996; Strout et al 1994).

Baker and Cynader (1994) recorded the responses of single units in cat visual cortex to stimuli consisting of F1–F2 motion sequences. In some sequences, the duration of F1 far exceeded the duration of the onset-locked transient response, yet strong motion signals were triggered when F2 was presented. Indeed, motion signals were produced even if F2 was presented while F1 was still on display, so that F1 and

¶ Author to whom correspondence and requests for reprints should be addressed.

F2 were on the screen together. Notably, at the time F2 was displayed, all the transient activity produced by F1 had abated, and only the ongoing sustained activity remained. The theoretical importance of these results was that they ruled out interaction between transient responses as the sole sources of motion signals. Instead, they suggested that motion signals could ensue from the interaction of transient and sustained responses, thus entailing the existence of a sustained input to the motion sensor.

On the strength of this result, models of motion sensors have been generalized to include not only transient but also sustained components (Baker and Cynader 1994; Bischof and Di Lollo 1996; Emerson et al 1992; Strout et al 1994; Wilson 1985). In such a sensor, a motion signal can be produced by the onset of F2 even after the transient signal produced by the onset of F1 has abated. The crucial requirement is for sufficient F1-related sustained activity to be still available at the onset of F2. In other words, a motion signal can be produced by the onset of F2 provided that F1 is still visible when F2 is presented.

1.1 *Related developments*

We obtained evidence consonant with this viewpoint in a series of experiments in which the stimulus configurations and the temporal parameters of the displays matched broadly the corresponding parameters in Baker and Cynader's (1994) experiments, notably the temporal overlap between F1 and F2. In our experiments (Bischof and Di Lollo 1996), all the display sequences included a period during which F1 and F2 were displayed concurrently. More specifically, all display sequences comprised two types of images, single images and combined images. Single images consisted of one set of dots, either F1 or F2, depending on the display sequence. The combined images contained both F1 and F2. All displays contained both types of images, but in different sequences, as illustrated in figure 1. In the 0-SOA (zero-stimulus-onset-asynchrony) sequence, the display began with the combined image and ended with a single image. In the 0-STA (zero-stimulus-termination-asynchrony) sequence, the display began with a single image and ended with the combined image. In the asynchronous (ASYN) sequence, the display began with F1, continued with the combined image, and ended with F2.

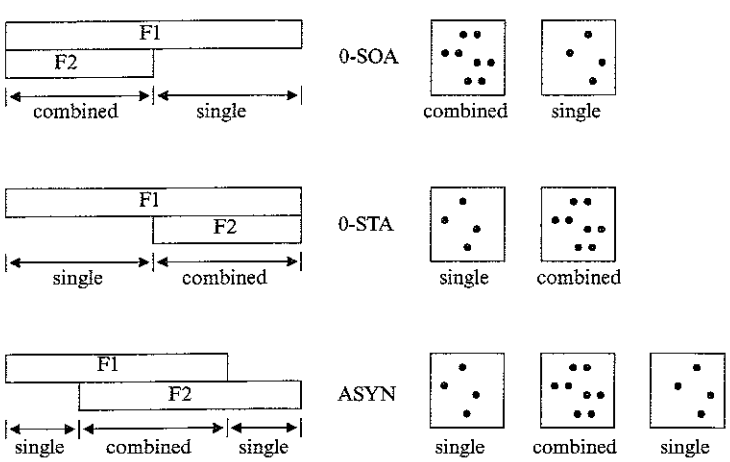

Figure 1. Schematic illustration of the temporal sequence of single and combined displays in the 0-SOA, 0-STA, and ASYN stimulus sequences. For each of the three stimulus sequences, the temporal course is exemplified on the left and a sketch of their spatial configuration is exemplified on the right. All configurations illustrated here produced the perception of rightward motion. Converse configurations were used to produce leftward motion. Note that the number of dots in the actual displays was far greater than shown in the sketches.

A modified version of Baker and Cynader's (1994) model, described in detail below, was supported by the human psychophysical data obtained with these stimuli (Bischof and Di Lollo 1996). In other words, the motion sensor in the new model could handle both transient and sustained inputs. The new model predicts that perception of directional motion should continue unabated, and should not decline, however long the SOA or STA. Empirical findings confirmed this prediction, but the evidence was not entirely uniform: there were suggestions that, under some viewing conditions, the strength of the motion signal may decline at long SOAs or STAs (Baker and Cynader 1994; Bischof and Di Lollo 1996). One way in which this could occur is suggested below in equation (4), which shows that strength of the motion signal is constrained by the magnitude of the ongoing visual activity triggered by F1. Were the magnitude of this activity to decline progressively after F1 onset, a corresponding reduction in strength of motion signal would follow. The present work is a test of this conjecture.

1.2 *Rationale of the new work*

We make the working assumption that the visual response to a step stimulus comprises both transient and sustained components. Such a response pattern has been observed in the activity of cortical neurons that show step responses consisting of an initial transient discharge followed by a much smaller sustained component (eg Baker and Cynader 1994). Further, it is known that the transient component is more dominant in high-luminance than in low-luminance viewing conditions. This relationship has been found in neurophysiological investigations (Hood and Grover 1974; Whitten and Brown 1973) and is consistent with psychophysical findings (Barlow 1958; Di Lollo and Bischof 1995; Roufs 1972; Watson 1986). Our earlier experiments were performed under low-luminance conditions; therefore, the transient component was probably quite weak and the sustained component relatively more dominant. In turn, the moderately high level of sustained activity present at the onset of F2 would have allowed relatively strong motion signals to be produced even at long SOAs or STAs. On this reasoning, a different outcome should be expected were the experiments to be done under high-luminance conditions: the relatively low level of sustained activity would reduce the input to the motion sensor at long SOAs or STAs, with consequent reduction in the strength of the motion signals. The principal objective of the present work was to examine this supposition by examining these effects under high-luminance viewing conditions. In pursuing this objective, we employed stimulus parameters and model parameters closely matched to those of our earlier work. In a second set of experiments we pursued some findings in the ASYN condition that did not match model predictions.

2 Experiment 1

2.1 *Methods*

Methods were essentially the same as in our earlier experiments (Bischof and Di Lollo 1996), with the exception that the adapting luminance was in the photopic range, as explained below.

2.1.1 *Observers.* The third author and a paid volunteer, both experienced observers, served in experiment 1. Both had corrected-to-normal vision, and had participated in an earlier experiment (Bischof and Di Lollo 1996) approximately two weeks earlier.

2.1.2 *Apparatus.* All stimuli were composed of dots displayed on a Tektronix 608 oscilloscope equipped with P15 phosphor. The X, Y, and Z (intensity) coordinates of each dot were stored in a fast plotting buffer that displayed them to the screen at the rate of 1 dot μs^{-1} (Finley 1985). Intensity of the dots was measured with a Minolta LS-110 luminance meter by the method described by Sperling (1971). Values of luminous directional energy were expressed in millicandela microsecond (mcd μs)—rather than the

more common candela microsecond—in order to avoid very small numbers. Images were displayed within a square area in the centre of the screen. At the viewing distance of 57 cm, set by a headrest, the square area subtended an angle of 2 deg. The screen was front illuminated with a Kodak Carousel projector equipped with a 500W GE Quartzline projector lamp. The light beam passed through two filters; the first was custom-made and blocked all wavelengths below about 410 nm and above about 725 nm. The second filter was a Kodak Wratten 38A filter with peak transmission at 440 nm. This filtering produced illumination with a spectral composition similar to that of the dots. The light beam was further attenuated by means of neutral-density filters to produce a screen luminance of 10 cd m^{-2}.

2.1.3 *Visual displays.* All displays comprised two sets of dots, F1 and F2. Set F1 contained 160 dots distributed randomly on each trial over the 2 deg × 2 deg viewing area. Set F2 also contained 160 dots; 80 were distributed randomly on each trial over the 2 deg × 2 deg viewing area; the other 80 were identical to 80 randomly chosen dots in F1 except for a uniform horizontal displacement of 10 min arc; that is, only 50% of the dots were displaced coherently. This was done to lower the level of performance and thus prevent excessive ceiling effects. Direction of displacement (left or right) was determined randomly on each trial. The displacement occurred entirely within the spatial confines of the 2 deg × 2 deg viewing area, with conventional wraparound for dots displaced outside the viewing area. The size of each dot was approximately 0.25 min arc.

Every motion sequence has two components: the single image and the combined image. The single image consisted of a single set of dots, either F1 or F2, depending on the display sequence. The combined image contained both F1 and F2, a total of 320 dots. There were three basic sequences, as illustrated in figure 1. In the 0-SOA sequence, the display began with the combined image and ended with a single image; in the 0-STA sequence, the display began with a single image and ended with the combined image; in the ASYN sequence, the display began with a single image (F1), continued with the combined image, and ended with a single image (F2).

There were six exposure durations of the single images and of the combined images: 10, 20, 40, 80, 160, and 320 ms. The six durations of the single images were combined factorially with the six durations of the combined images to yield a total of 36 conditions for each of the 3 basic sequences illustrated in figure 1, for a total of 108 temporal sequences. It should be made clear that, in the ASYN condition, where the sequence began and ended with a single image (see figure 1), each single image was displayed for the entire pertinent duration. For example, for single and combined durations of 40 and 320 ms, respectively, the total duration of the display in the ASYN condition was 400 ms.

Intensity of the dots. For all conditions in which the durations of F1 and F2 differed, an important consideration was time–intensity reciprocity known as Bloch's law. That is, up to a critical duration, the brightness (ie the perceived intensity) of a stimulus is determined jointly by its luminance and its duration. Such time–intensity reciprocity holds for intensities near threshold (Bloch's law) as well as above threshold (Butler 1975). Because of time–intensity reciprocity, increasing the exposure duration of a stimulus would result in a corresponding increment in brightness. Thus, the effect of duration would be inextricably confounded with that of brightness. In earlier investigations, this confounding was resolved by displaying briefer stimuli at higher luminance values, so that all stimuli appeared of the same brightness regardless of duration (Bischof and Di Lollo 1996; Bowen et al 1974; Servière et al 1977). This approach yields equal-energy stimuli up to the critical duration, and was implemented in the present work. The process of brightness equalization was carried out in a separate psychophysical procedure similar to those described by Servière et al (1977) and by Di Lollo and Finley (1986).

In this procedure, two identical random-dot stimuli are displayed briefly side by side for different durations (eg 20 and 30 ms, respectively). When the two are displayed at the same level of luminance, the longer stimulus appears brighter because of Bloch's law. The task of the observer is to adjust (reduce) the luminance of the longer stimulus until its brightness matches that of the shorter stimulus. The shortest stimuli (10 ms) were displayed at a luminous directional energy of 13.4 mcd μs; longer stimuli were displayed at correspondingly lower levels of intensity. This matching procedure was carried out individually by each observer, and the displays were compensated individually. The results of the compensation procedure for observer VDL are shown in figure 2; virtually identical results were obtained for observer RH.

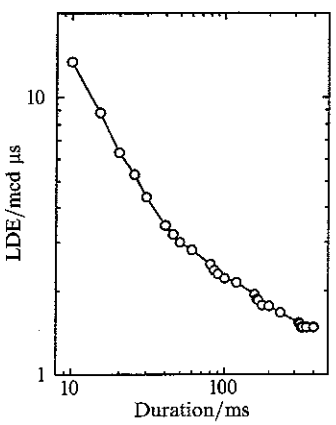

Figure 2. Luminous directional energy (LDE) to which the observers set a comparison stimulus displayed for the duration shown on the abscissa, so that it appeared of the same brightness as a standard stimulus displayed for 10 ms at an LDE of 13.4 mcd μs. Each point defines a combination of intensity and duration that yields the same brightness as any other point.

2.1.4 *Procedure.* Observers sat in a dimly lit observation chamber. On each trial, the sequence of events was as follows. A dim fixation cross (0.5 deg × 0.5 deg) was displayed in the centre of the screen. When ready, the observer pressed a push button on a hand-held box to initiate a trial. Upon a button press, the fixation cross disappeared and, after a delay of 200 ms, the appropriate sequence of stimuli was displayed. Next, the observer reported (or guessed) the direction of motion (left or right) by pressing the appropriate button on the box. Finally, the fixation cross reappeared to indicate readiness for the next trial. An experimental session lasted about one hour. Each observer made 100 observations in each of the 108 experimental conditions.

2.2 *Results*

Individual results are shown in figure 3, separately for each of the three motion sequences. In every case, accuracy in identifying the direction of motion increased as the duration of the single image was increased from very brief to intermediate levels, and then decreased at longer durations. In other words, perception of directional motion was optimal at intermediate values of SOA or STA, but it declined at higher values. The inverted-U-shaped curves in figure 3 differ sharply from the monotonic ascending curves obtained under low-luminance viewing conditions (Bischof and Di Lollo 1996).

In contrast to the nonmonotonic changes in performance as a function of the duration of the single component (figure 3), performance decreased monotonically as the duration of the combined image was increased. The monotonic decrement can be seen in figure 3 by noting the order of the curves within each panel. In this respect, performance as a function of duration of the combined image follows similar trends in low-luminance to those in high-luminance viewing conditions.

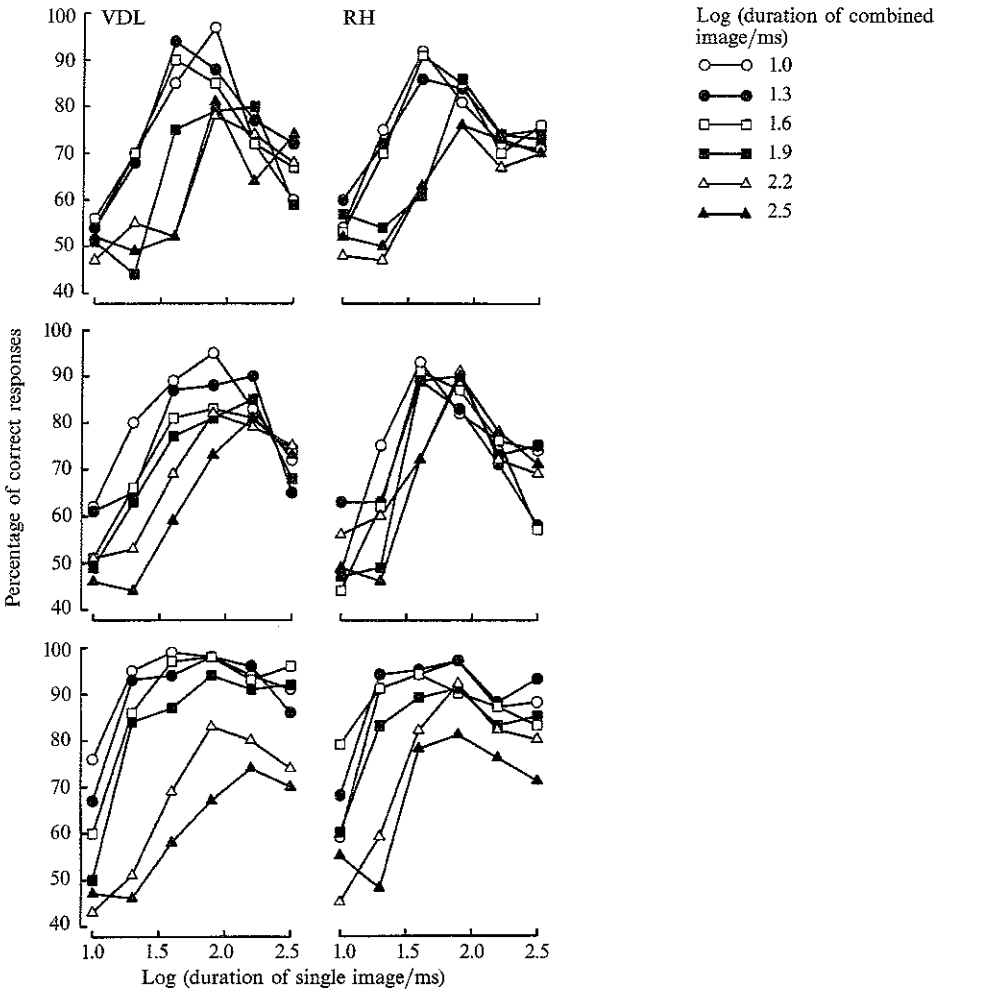

Figure 3. Illustrated within each panel are the results for each of the 36 experimental conditions resulting from the factorial combination of six durations of the single image and six durations of the combined image. Results for the two observers (VDL and RH) are shown in separate panels for each of the three stimulus sequences (top, 0-SOA; middle, 0-STA; bottom, ASYN).

2.3 *Discussion*

2.3.1 *A modified model.* Direct application of Bischof and Di Lollo's (1996) model failed to provide an acceptable account of the results obtained in the present high-luminance conditions. Below, we present a modified model that provides suitable accounts of the outcomes obtained in both low-luminance and high-luminance viewing. In the earlier model, a model signal was generated through an interaction between a fast transient input (T) and a slow sustained input (S). In the current model, T remains unchanged, but S is modified to include both a sustained (S_S) and a transient (S_T) component. The relative prominence of S_S and S_T is assumed to depend on level of adaptation, with the prominence of S_T increasing with light adaptation, as specified in the equations below. We should note that this is one way in which the luminance dependence of the sustained input can be modelled. Alternative formulations have been used by Emerson et al (1992) and by Strout et al (1994).

In the present model, the input is convolved with a pair of temporal filters, T and S, at each spatial position. The T-filters are assumed to generate transient outputs, whereas

S-filters are assumed to generate outputs containing both transient and sustained components. More precisely, the transient filter T, taken from Adelson and Bergen (1985), is defined by the filter equation

$$T(t; n) = \left(\frac{t}{\tau}\right)^n \exp\left(\frac{-t}{\tau}\right)\left[\frac{1}{n!} - \frac{t^2}{\tau^2(n+2)!}\right], \tag{1}$$

with $n = 3$ and $\tau = 10$ ms. The S-filter is defined by a mixture of transient and sustained components, $S_T(t)$ and $S_S(t)$, with the transience parameter α defining the relative weight of the two components:

$$S(t) = \frac{\alpha}{M_T} S_T(t) + \frac{1-\alpha}{M_S} S_S(t), \tag{2}$$

where $S_T(t) = T(t; 5)$, $M_T = \max S_T(t)$, $M_S = \max S_S(t)$, and

$$S_S(t) = \int T(t; 3)\mathrm{d}t = \left(\frac{t^5}{5! \cdot \tau^4} + \frac{t^4}{4! \cdot \tau^3}\right) \exp\left(\frac{-t}{\tau}\right), \tag{3}$$

again with $\tau = 10$ ms.

The impulse response function of the T-filter is illustrated in figure 4a along with those of the S-filter for two values of α. Figure 4b illustrates the corresponding responses to a pulse of 160 ms.

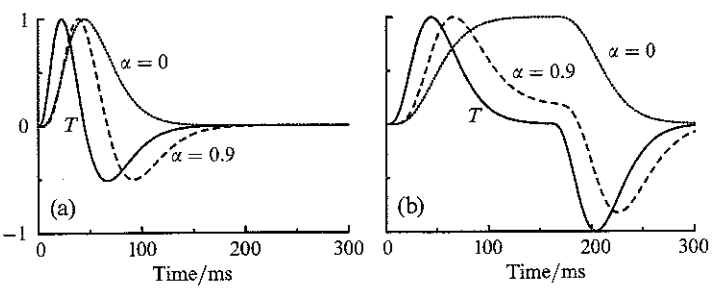

Figure 4. (a) Impulse response functions of the T-filter (continuous line), of the S-filter with $\alpha = 0.9$ (segmented line), and of the S-filter with $\alpha = 0$ (dotted line); (b) corresponding responses to a pulse of 160 ms. All functions have been normalized to unity.

The new model is summarized in figure 5. The central portion of figure 5 shows the structure of the motion sensor. Each surrounding graph illustrates the temporal course of activity in the corresponding part of the sensor over an interval of 400 ms from the onset of F1. Exemplified in figure 5 is a 0-STA sequence with a 40 ms single image and an 80 ms combined image. Two sequential stimuli, F1 and F2, are assumed to activate the motion sensor through spatially separate inputs I_L and I_R, respectively. The outputs of the salient filters are combined (as described below) in C_L and C_R, with C_L responding optimally to motion in the direction $I_R - I_L$ and C_R responding optimally to motion in the direction $I_L - I_R$. As was done by Baker and Cynader (1994), the combination of the signals in C_L and C_R is modelled in heterosynaptic facilitation and can be represented as

$$C(t) = T(t)[1 + S(t)]. \tag{4}$$

The directional motion signal is given by the difference $C_R - C_L$. The magnitude of the motion signal is normalized as follows:

$$\Sigma_N(t) = \frac{C_R(t) - C_L(t)}{|C_R(t)| + |C_L(t)|}. \tag{5}$$

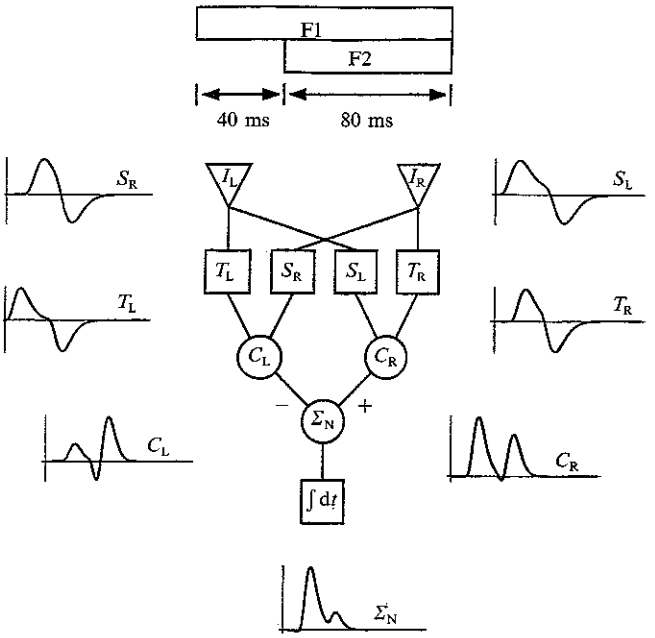

Figure 5. Schematic illustration of a modified Reichardt motion sensor. Input to the sensor passes through a spatial filter (I_L or I_R) and a pair of temporal filters: one generates transient responses (T_L or T_R), the other generates responses containing both transient and sustained components (S_L or S_R). Two correlators (C_L and C_R) combine the outputs from the temporal filters. The magnitude of the motion signal (Σ_N) is the normalized difference between the responses of the two correlators. Responses by each component of the model over a temporal interval of 400 ms are shown for the 0-STA example with a 40 ms single image and an 80 ms combined image. Positive, negative, or zero sums indicate motion to the right, to the left, or no motion, respectively.

Normalization, as given in equation (5), is required to account for all the present results. This approach to normalization is closely related to Heeger's (1992) divisive normalization and to Geisler and Albrecht's (1992) multiplicative contrast-gain control mechanisms. Our equation is simpler than either Heeger's or Geisler and Albrecht's because we did not take into account contrast-response and saturation effects. We could omit these factors because our stimuli were far above threshold, and it is known that the motion system saturates rapidly at low contrast levels.

2.3.2 *Application of the model to the data.* Application of the model to psychophysical data is predicated on the assumption that the strength of the motion signal, integrated over time as in figure 5, is a predictor of performance in the motion-detection task. In our simulation, we aimed at providing only qualitative accounts of the experimental outcomes, as distinct from the quantitative fits of the data, and we considered only average levels of performance in each of the three experimental conditions illustrated in figure 3.

Average performance curves for observer VDL are shown in figures 6, 7, and 8 for conditions 0-SOA, 0-STA, and ASYN, respectively. Performance averaged over all durations of the combined image is shown in figures 6a, 7a, and 8a. The corresponding results obtained in an earlier study by the same observer in low-luminance viewing (see Bischof and Di Lollo 1996) are also shown in each figure to permit direct comparison with the present results. The performance curves of observer RH are not shown because they were very similar to those of VDL. Predicted scores, averaged over all durations of the combined image, separately for low-luminance and high-luminance

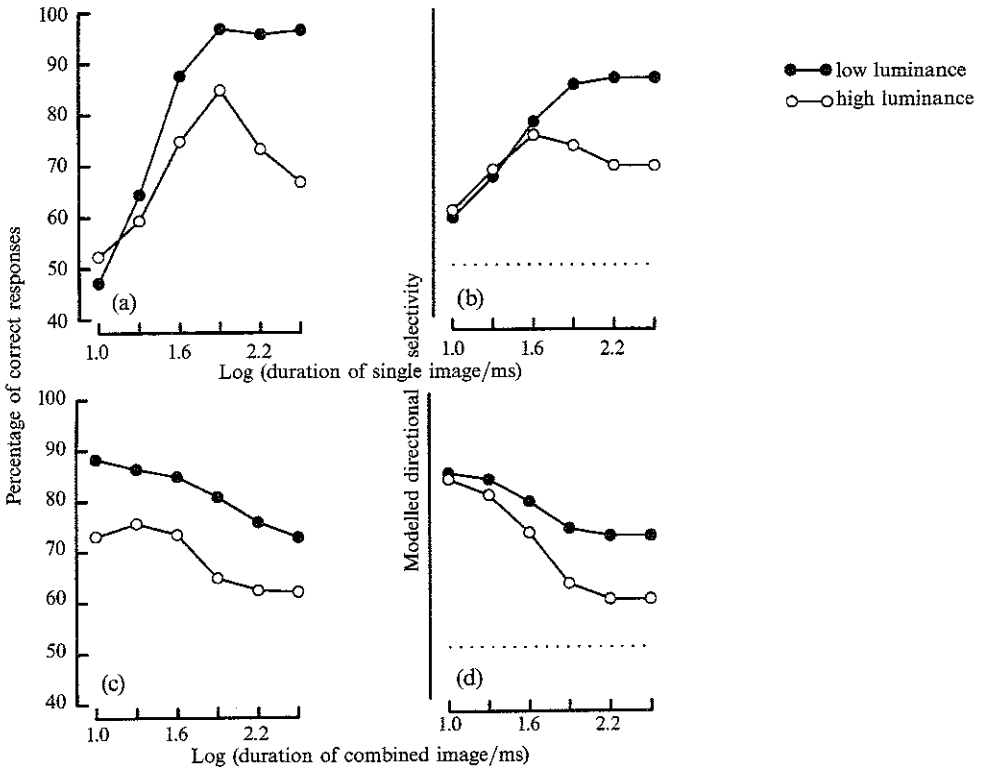

Figure 6. Obtained (observer VDL) and simulated results with the 0-SOA stimulus sequence. (a) Percentage of correct responses at each duration of the single image, averaged over all durations of the combined image. (c) Percentage of correct responses at each duration of the combined image, averaged over all durations of the single image. In both graphs, the open circles are the results from the high-luminance viewing condition (experiment 1) and the filled circles are the results from the low-luminance viewing condition (see Bischof and Di Lollo 1996) for the same observer. (b) and (d) correspond to (a) and (c), respectively, and show simulated results from the model described in the text and illustrated in figure 5. The ordinates are scaled in arbitrary units; the same scaling factor was used in both panels. The dotted lines indicate the zero level in modelled directional selectivity.

viewing, are shown in figures 6b, 7b, and 8b, for conditions 0-SOA, 0-STA, and ASYN, respectively. Corresponding performance curves and simulated scores, averaged over all durations of the single image, are shown in panels (c) and (d), respectively, in figures 6, 7, and 8. In the simulations, a different arbitrary scaling factor was used for the low-luminance and high-luminance conditions and for each stimulus sequence (0-SOA, 0-STA, ASYN); however, the same scaling factor was used for each combination of stimulus sequence and luminance [ie corresponding curves in panels (b) and (d) in each figure had the same scaling factor]. All other parameters in the low-luminance simulations were the same as in the corresponding high-luminance simulations, except for the transience parameter α which was set to 0.9 in the light and to zero in the dark. Setting α to zero produced simulated S_L and S_R responses that were entirely sustained, and rendered the present low-luminance simulation equivalent to that reported by Bischof and Di Lollo (1996).

2.3.3 *Empirical verification of the model*

Duration of the single images. Figures 6a, 7a, and 8a show mean performance as a function of the duration of the single image. Within each panel, the curves reveal pronounced differences in both level and shape between low-luminance and

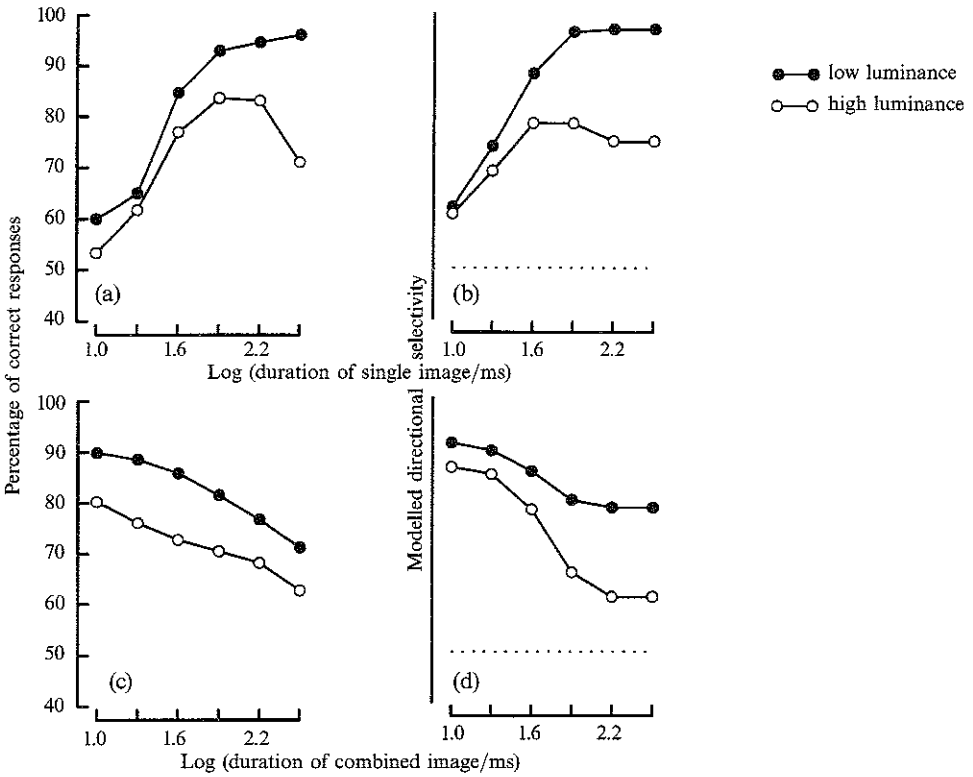

Figure 7. Same as figure 6, for the 0-STA stimulus sequence.

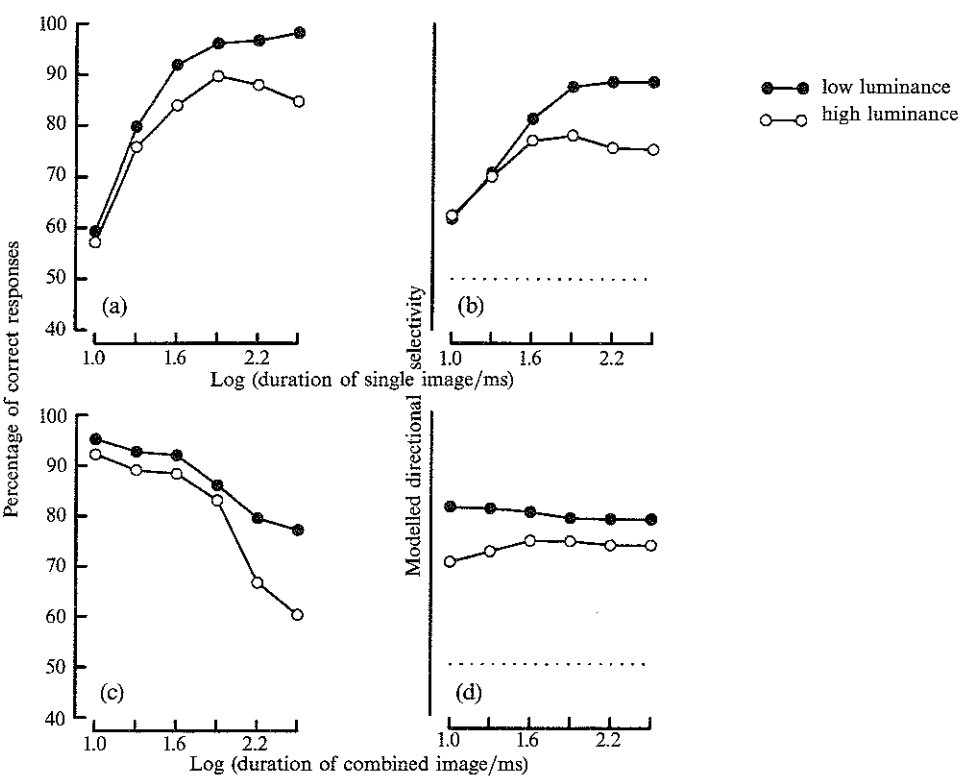

Figure 8. Same as figure 6, for the ASYN stimulus sequence.

high-luminance conditions. Of particular interest are the downturns exhibited at the longer durations of the single image in the high-luminance but not in the low-luminance conditions. These differences are reproduced quite well in the simulated curves, although there are some differences of detail. For example, predicted peak performance as a function of duration of the single image in the high-luminance condition (figures 6b, 7b, and 8b) is predicted to occur at about 40 ms. Instead, the obtained peak occurred at about 80 ms (figures 6a, 7a, and 8a). In this respect it should be stressed that it was not our intent to provide a quantitative model of the results, therefore we did not employ any parameter-fitting procedures. Rather, our intent was to provide a qualitative model that captured the overall trends in the data.

Even though illustrative, the simulations do not provide an immediately intuitive description of the underlying processing events. A descriptive account of the downturns in performance can be given in terms of the motion signals triggered by the interaction between the F2 response and the ongoing F1 response. The latter consists of the sum of transient (S_T) and sustained (S_S) components. The sustained component remains at the same level throughout the period for which the stimulus is visible; by contrast, the transient component first rises to a maximum and then declines. In turn, the level of total activity rises to a maximum with the F1 transient response and then declines to the level of F1-sustained activity. For any given F2, the strength of the motion signal will increase with the level of ongoing F1 activity. Thus, in terms of the model, the decline seen in the high-luminance curves represents the decline of the transient component in the F1 response. In the simulation, the decline in activity is controlled by the value of the transiency parameter α, which varies betwen zero (for responses with no transient component) and unity (for responses with no sustained component).

Duration of the combined image. Figures 6c, 7c, and 8c show mean performance as a function of the duration of the average combined image. The corresponding predictions are shown in figures 6d, 7d, and 8d. The simulated curves provide good qualitative matches to the empirical curves, except for the ASYN condition, which is discussed further below. Overall, the present model is capable of accounting for the outcomes obtained both with high luminance, as in the present study, and with low luminance, as in our earlier study (Bischof and Di Lollo 1996). An exception to the success of the model is seen in the simulated ASYN performance as a function of the duration of the combined image. The steadily declining empirical curves in figure 8c are predicted in figure 8d as essentially flat, if not actually increasing. This failure is surprising because the ASYN sequence can be regarded as a composite of two sequences (0-STA followed immediately by 0-SOA) that, taken separately, create no problem for the model. Obviously, the ASYN sequence must differ from the 0-SOA and 0-STA sequences in ways that go beyond simple summation.

One way in which the ASYN sequence differs from the other two is in respect to the number of motion signals that are generated. The 0-STA and 0-SOA sequences produce one signal each. The ASYN sequence produces two motion signals: a leading signal, as in a 0-STA sequence, and a trailing signal, as in a 0-SOA sequence. In experiment 1, the temporal separation between the two signals was varied over a considerable range. The effect of separation, however, could not be assessed unambiguously because it covaried with the duration of the combined image. Therefore, the two factors were confounded as determinants of motion detectability. Experiment 2 was designed to resolve this confounding.

3 Experiment 2

3.1 *Methods*

In experiment 2, our objective was to construct a set of stimuli capable of triggering the same motion signals as the corresponding ASYN stimuli in experiment 1, while

avoiding a confounding with the duration of the combined image. This was achieved by replacing part of the combined image with a temporal gap during which no stimuli were displayed and the background luminance of the screen remained unchanged. Each stimulus consisted of an initial 0-STA component followed by a 0-SOA component. The duration of the single image in each component was the same as that in the corresponding stimulus in experiment 1 (10, 20, 40, 80, 160, or 320 ms). The duration of the combined image was fixed at 10 ms for each component. We employed a factorial design in which the six durations of the single image were crossed with six durations of the 'mid-stimulus interval' (MSI): 10, 20, 40, 80, 160, or 320 ms. The MSI corresponded to the duration of the combined image in experiment 1, and was defined as the interval from the onset of the leading combined image (in the 0-STA component) to the offset of the trailing combined image (in the 0-SOA component), including the temporal gap, if required. The factorial combination of MSI and duration of the single image yielded 36 stimuli, as exemplified in figure 9.

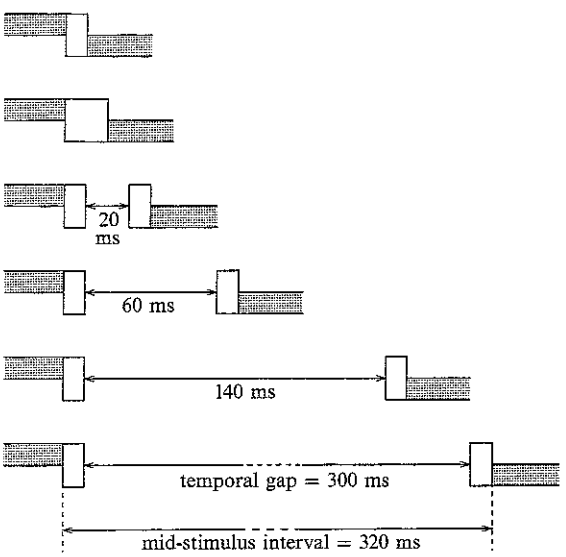

Figure 9. Schematic illustration of the conditions used in experiment 2. Each stimulus consisted of an initial 0-STA component followed by a 0-SOA component. The duration of the combined image was fixed at 10 ms for each component. The mid-stimulus interval (MSI) denotes the temporal interval between the two single images; it included both combined images and a temporal gap, when required. The durations used for the MSIs were the same as those used for the combined image in experiment 1. Thus, in the four longest MSIs, there was a temporal gap of 20, 60, 140, or 300 ms, as shown.

Stimuli were displayed at two levels of adapting luminance. The high adapting luminance was the same as in experiment 1; the screen was front illuminated to 10 cd m^{-2}, and the intensity of the dots was equalized as in experiment 1. The low adapting luminance (less than 0.1 cd m^{-2}) was the same as that employed by Bischof and Di Lollo (1996): the ambient illumination was dim, the display screen was dark, and the intensity of the dots was equalized by the same procedure as in experiment 1. In all other respects, the procedures were the same as in experiment 1.

3.2 *Results and discussion*

The results of the two observers are shown in figure 10, separately for the two levels of adapting luminance. Overall, the curves exhibit an inverted-U shape similar to that obtained in experiment 1 (figure 3). An interesting pattern is revealed by the order of the curves within each panel: the lowest performance was often obtained not at

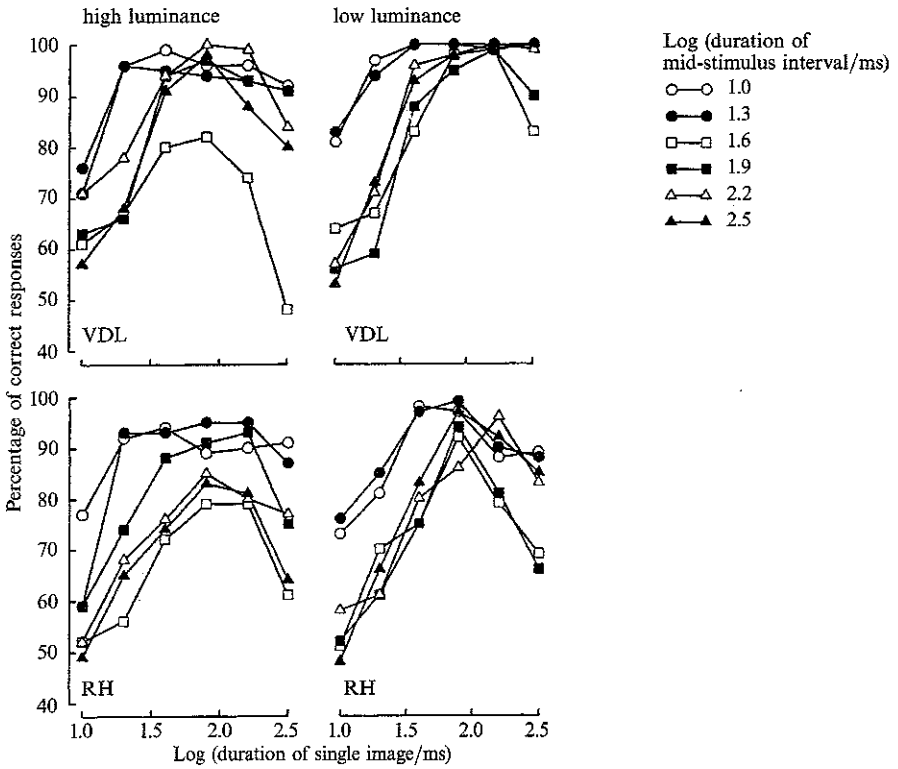

Figure 10. Obtained results from experiment 2. Separate panels show results for the two observers under two viewing conditions. Illustrated within each panel are the results for each of the 36 experimental conditions resulting from the factorial combination of the six durations of the single image and the six durations of the mid-stimulus interval.

the longest, but at the intermediate, durations of MSI. This is particularly evident at the longer durations of the single image. Observer VDL's performance, averaged over all MSIs, is shown in figure 11a, separately for the two levels of adapting luminance. Performance averaged over all durations of the single image is shown in figure 11c. A notable aspect of the high-luminance curve in figure 11c is the dip in performance at the 40 ms MSI, followed by a recovery at the longer MSIs. The dip corresponds to the low performance at the intermediate MSIs seen in figure 10. The effect is shown most clearly in the results of observer VDL, but the same pattern is exhibited by observer RH.

Predictions from the model were derived by the procedures described in experiment 1. The main results of the simulation are shown in figures 11b and 11d. The pattern of predicted outcomes is similar to that for experiment 1 [panels (b) and (d) in figures 6, 7, and 8], with one notable exception. The predicted curve for the high-luminance condition in figure 11d exhibits a sharp decrement at the intermediate durations of MSI (corresponding to temporal gaps of 20 and 40 ms), followed by a recovery at the longer durations. This matches the dip in the corresponding empirical curve in figure 11c, thus confirming predictions from the model.

Overall, the theoretical predictions provided good qualitative matches to the empirical data. The results attest to the ability of the model to cope with sequential motion signals separated by temporal gaps. In this respect, the mismatch between empirical and predicted curves seen in figures 8c and 8d must be attributed not to the occurrence of two sequential motion signals, but to as yet unexplained effects of long-lasting combined images in the ASYN condition.

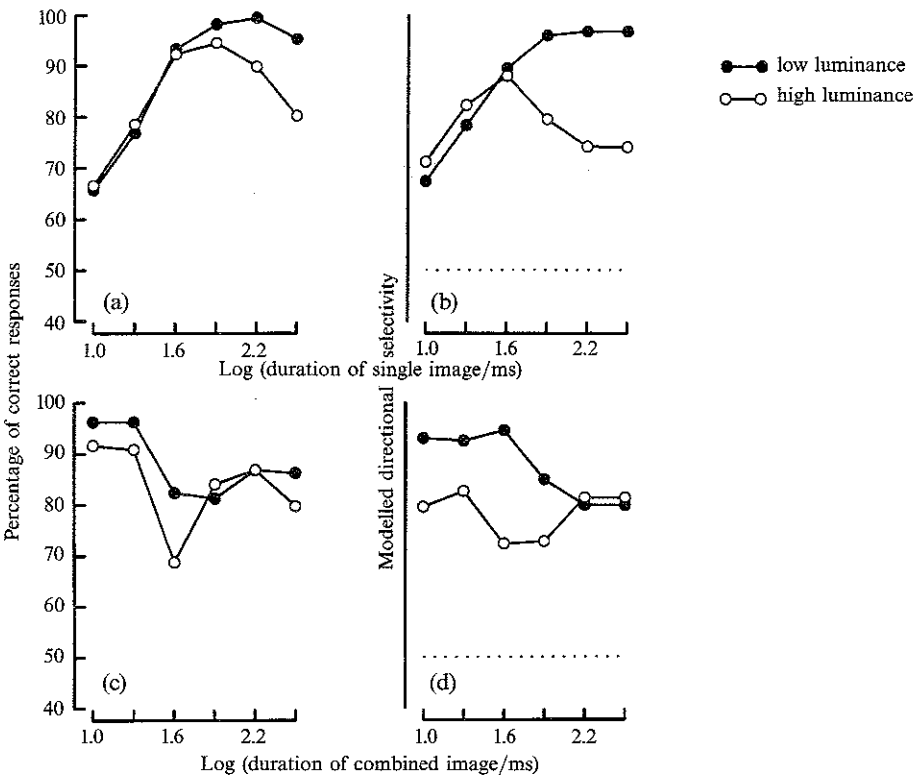

Figure 11. Obtained (observer VDL) and simulated results for experiment 2. (a) Performance averaged over all mid-stimulus intervals (MSIs) for each duration of the single image. (c) Performance averaged over all durations of the single image for each MSI. In both panels, the open circles are the results from the high-luminance viewing condition, and the filled circles are the results from the low-luminance viewing condition. (b) and (d) correspond to (a) and (c), respectively, and show simulated results. The dotted lines indicate the zero level in modelled directional selectivity.

It is interesting to note that reverse motion (ie motion in the direction opposite that of the spatial displacement) was never observed in the present work. In contrast, strong reverse motion was obtained by Strout et al (1994) when the space-average luminance of the screen did not change during the interframe interval, which corresponded to the temporal gap in the present displays. The different outcomes cannot be ascribed to luminance factors because the dots in our displays did not contribute significantly to overall luminance. Therefore, it can be assumed that screen luminance remained essentially constant during the temporal gap in both studies. On the other hand, the different outcomes might be attributable to differences in stimulus configurations. In the experiment of Strout et al (1994), the stimulus configuration changed across the temporal gap. That is, the stimulus displayed after the gap was a translated version of the pregap stimulus. In contrast, in our experiment, the pregap and postgap stimulus configurations were identical (see figure 9). This issue cannot be resolved without further studies expressly designed for this purpose. The performance with sequential motion signals separated by temporal gaps is studied further in experiment 3.

4 Experiment 3

In experiment 2, an unanticipated dip in performance was obtained at a relatively brief duration of the temporal gap. In that experiment, duration of the temporal gap was varied over a substantial range, while the duration of the combined image was held

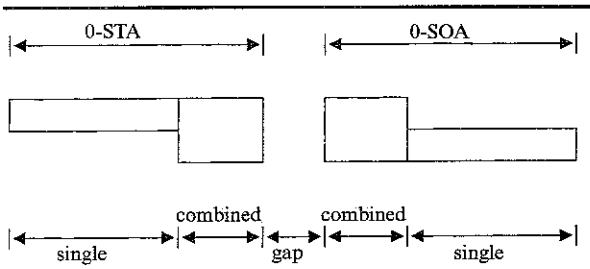

Figure 12. Schematic illustration of the sequence of events in experiment 3. Duration of the single image was fixed at 320 ms. Three durations of the combined image (10, 20, or 40 ms) were used with each of six durations of the temporal gap (0, 10, 20, 40, 80, or 160 ms).

constant and brief. To explore the determinants of the performance dip, we investigated parametrically the predictions of the model for a range of durations of the temporal gap and of the combined image. The simulations showed that the dip in performance was obtained only at very brief durations of the combined image. More specifically, the dip was obtained with combined images of 10 and 20 ms, but not with combined images of 40 ms or beyond. Experiment 3 was a verification of these predictions.

Each stimulus comprised an initial 0-STA component followed by a 0-SOA component, as illustrated in figure 12. The duration of the single image was fixed at 320 ms. We employed a factorial design in which three durations of the combined image (10, 20, or 40 ms) were crossed with six durations of the temporal gap (0, 10, 20, 40, 80, or 160 ms). The stimuli were displayed at two levels of adapting luminance: low or high, as defined in experiment 2. In all other respects, the procedures were the same as in experiment 2.

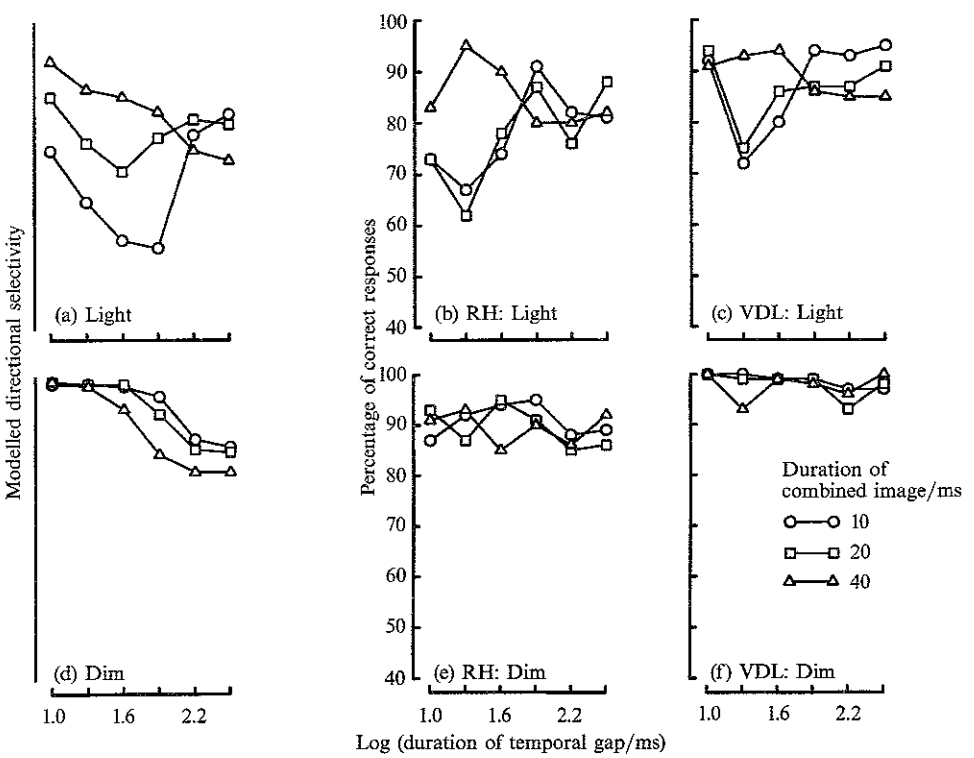

Figure 13. Simulated and obtained results from experiment 3. (a) and (d) Predictions from the model for the high-luminance and low-luminance viewing conditions, respectively; (b) and (e) correspond to (a) and (d), respectively, and show results from observer RH; (c) and (f) correspond to (a) and (d), respectively, and show obtained results from observer VDL.

Predictions from the model, derived as in experiments 1 and 2, are illustrated in figures 13a and 13d. Empirical results for the two observers are shown in figures 13b and 13c for high-luminance viewing, and in figures 13e and 13f for low-luminance viewing. Once again, the simulation provides a good qualitative prediction of the empirical outcomes. In high-luminance viewing, the simulated curves in figure 13a reveal a dip in performance with combined images of 10 and 20 ms, but not of 40 ms. The dips in performance are reflected remarkably well in the empirical curves in figures 13b and 13c. By the same token, a dip in performance was not obtained by either observer in the low-luminance viewing condition (figures 13e and 13f), reflecting the absence of a dip in the corresponding simulated curves. A discrepancy may be noted between empirical and simulated curves in low-luminance viewing: the empirical curves do not reflect the downward trend exhibited by the simulated curves. It is possible, however, that the decline occurred but could not be seen because the performance levels shown in figures 13e and 13f were close to ceiling. Lowering the overall level of performance—perhaps by lowering the signal-to-noise ratio (ie by reducing the proportion of coherently moving dots below the 50% used in this experiment)—might permit the decline to be seen.

5 Concluding remarks

The present work is based on the model of motion perception illustrated in figure 5. In that model, a motion signal is generated through an interaction between a fast transient input and a slower, more sustained input. The temporal characteristics of the latter are deemed to depend on adaptation level: the input is completely sustained at low luminance, but it develops increasingly transient characteristics as luminance is increased. This relationship with adapting luminance is consistent with earlier findings that revealed an inverse relationship between level of adaptation and the temporal characteristics of motion perception (Bonnet 1982; Dawson and Di Lollo 1990; Morgan and Ward 1980; van Nes et al 1967). In the present work, we make the additional restrictive assumption that adapting luminance affects the slow but not the fast input to the motion sensor.

Predictions from the model were confirmed with both conventional and unconventional motion sequences (0-SOA, 0-STA, ASYN). In addition, the model generated nonintuitive predictions that were confirmed in experiments 2 and 3. To our knowledge, those effects had never been previously reported.

Formally, the present model bears distinct similarities to that proposed by Emerson et al (1992). Both models employ temporal impulse-response functions with transient and sustained components as inputs to directionally selective motion sensors. One of the major differences between the two models lies in our inclusion of normalization, as set out in equation (5). Clear parallels also exist between our model and that proposed by Strout et al (1994). Except for normalization, our model is similar to Strout et al's 'Phase II' which, in turn, is very similar to the model of Emerson et al (1992). A second important difference between our model and that of Strout et al regards the inputs to the motion sensor. Both models postulate two separate inputs to the motion sensor. Differences arise, however, in how the two input signals are deemed to vary in degree of transiency. In Strout et al's model, transiency is deemed to vary *concomitantly* in the two signals. By contrast, we assume that the transiency of one input remains constant, while the other varies with adapting luminance.

We performed a large set of simulations to check whether the models of Emerson et al (1992) and of Strout et al (1994) could account for the present outcomes. In all simulations, the transiency of both the fast and the slow signals were varied concomitantly by varying the value of B [see equation (2) in Strout et al 1994] from 0 to 1, in steps of 0.1. Without normalization, predictions from both models were adequate, except for the

effect of the combined image which was predicted as an inverted-U function of duration instead of a monotonically decreasing function. With normalization, on the other hand, the model of Strout et al (1994) provided adequate accounts of the 0-SOA and ASYN results, but predicted reverse motion for several 0-STA sequences in both high-luminance and low-luminance viewing. As noted earlier, reverse motion was never observed with our stimulus sequences. Normalization, as expressed in equation (5), can be implemented easily in motion-opponent models (eg Strout et al 1994). In contrast, while normalization can be added to motion-energy models (Heeger 1992), it cannot be implemented exactly as expressed in equation (5), as can be seen in figures 7b and 7c in Emerson et al (1992). We ascribe the more accurate predictions of our model to two factors: implementation of normalization, and allowance for the transiency of the two input signals to vary independently—rather than concomitantly—in response to changes in adapting luminance.

Acknowledgements. This work was supported by grants OGP38521 and OGP0006592 from the Natural Sciences and Engineering Research Council of Canada to Walter F Bischof and to Vincent Di Lollo, respectively.

References

Adelson E H, Bergen J R, 1985 "Spatiotemporal energy models for the perception of motion" *Journal of the Optical Society of America A* **2** 284–299

Baker C L Jr, Cynader M S, 1994 "A sustained input to the direction-selective mechanism in cat striate cortex neurons" *Visual Neuroscience* **11** 1083–1092

Barlow H B, 1958 "Temporal and spatial summation in human vision at different background intensities" *Journal of Physiology (London)* **141** 337–350

Bischof W F, Di Lollo V, 1996 "Psychophysical evidence of a sustained input to directionally selective motion mechanisms" *Perception* **25** 65–76

Bonnet C, 1982 "Thresholds of motion perception", in *Tutorials in Motion Perception* Eds A W Wertheim, W A Wagenaar, H W Leibowitz (New York: Plenum Press) pp 41–79

Bowen R W, Pola J, Matin L, 1974 "Visual persistence: Effects of flash luminance duration and energy" *Vision Research* **14** 295–303

Butler T W, 1975 "Luminance-duration relationships in the photopic ERG and the apparent brightness of flashes" *Vision Research* **15** 693–698

Dawson M, Di Lollo V, 1990 "Effects of adapting luminance and stimulus contrast on the temporal and spatial limits of short-range motion" *Vision Research* **30** 415–429

Di Lollo V, Bischof W F, 1995 "Inverse-intensity effect in duration of visible persistence" *Psychological Bulletin* **118** 223–237

Di Lollo V, Finley G, 1986 "Equating the brightness of brief visual stimuli of unequal durations" *Behavior Research Methods, Instruments, and Computers* **18** 582–586

Emerson R C, Bergen J R, Adelson E H, 1992 "Directionally selective complex cells and the computation of motion energy in cat visual cortex" *Vision Research* **32** 203–218

Finley G, 1985 "A high-speed point plotter for vision research" *Vision Research* **25** 1993–1997

Geisler W S, Albrecht D G, 1992 "Cortical neurons: isolation of contrast gain control" *Vision Research* **32** 1409–1410

Heeger D J, 1992 "Normalization of cell responses in cat retinal ganglion cells to brief flashes of light" *Visual Neuroscience* **9** 181–197

Hood D C, Grover B G, 1974 "Temporal summation of light by a vertebrate visual receptor" *Science* **184** 1003–1005

Morgan M J, Ward R, 1980 "Conditions for motion flow in dynamic visual noise" *Vision Research* **20** 431–435

Nes F L van, Koenderink J J, Nas H, Bouman M A, 1967 "Spatiotemporal modulation transfer in the human eye" *Journal of the Optical Society of America* **57** 1082–1088

Reichardt W, 1961 "Autocorrelation, a principle for the evaluation of sensory information", in *Sensory Communication* Ed. W A Rosenblith (Cambridge, MA: MIT Press) pp 303–317

Roufs J A J, 1972 "Dynamic properties of vision: I. Experimental relationships between flicker and flash thresholds" *Vision Research* **12** 261–278

Santen J P H van, Sperling G, 1985 "Elaborated Reichardt detectors" *Journal of the Optical Society of America A* **2** 300–320

Servière J, Miceli D, Galifret Y, 1977 "A psychophysical study of the visual perception of 'instantaneous' and 'durable' " *Vision Research* **17** 57–63

Sperling G, 1971 "The description and luminous calibration of cathode ray oscilloscope visual displays" *Behavior Research Methods and Instrumentation* **3** 148 – 151

Strout J J, Pantle A, Mills S L, 1994 "An energy model of interframe interval effects in single-step apparent motion" *Vision Research* **34** 3223 – 3240

Watson A B, 1986 "Temporal sensitivity", in *Handbook of Perception and Human Performance* Eds K R Boff, L Kaufman, J P Thomas (New York: John Wiley) pp 6-1 – 6-43

Watson A B, Ahumada A J, 1985 "Model of human motion-vision sensing" *Journal of the Optical Society of America* **24** 322 – 341

Whitten D N, Brown K T, 1973 "The time courses of late receptor potentials from monkey cones and rods" *Vision Research* **13** 107 – 135

Wilson H R, 1985 "A model for direction selectivity in threshold motion perception" *Biological Cybernetics* **51** 213 – 222