MOTION MECHANISMS HAVE ONLY LIMITED ACCESS TO FORM INFORMATION

Jonathan D. Victor and Mary M. Conte

Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue
and Laboratory of Biophysics, The Rockefeller University, 1230 York Avenue, New York City,
NY 10021, U.S.A.

Abstract—We investigate the roles of spatial frequency content, flicker and higher-order elements of form ("features") in the generation of motion percepts. These cues are separated through the use of dynamic visual stimuli based on stochastic textures. Flicker alone and spatial frequency content alone suffice to generate a strong motion percept, but higher-order elements of form alone generate a much weaker motion percept. Thus, even for achromatic stimuli, all pattern information is not equally available for motion processing. Furthermore, higher-order form information, which by itself does not provide a strong cue to motion, is shown to interact with other visual information to facilitate determination of direction of motion.

Human   Visual textures   Motion perception   Nonlinear interactions   Parallel pathways

INTRODUCTION

Physiological (Zeki, 1973, 1977; Hubel & Livingstone, 1987; Van Essen & Maunsell, 1983; Felleman & Van Essen, 1987) anatomical (Shipp & Zeki, 1985; Van Essen, Newsome, Maunsell & Bixby, 1986; Livingstone & Hubel, 1987a) and psychophysical (Livingstone & Hubel, 1987b; Treisman, 1982) evidence indicates that visual information processing is divided among several submodalities, including form, color, motion and depth.

Anatomically (Van Essen & Maunsell, 1983), central visual areas may be grouped into two sets: a parietal stream, which contains neurons with large receptive fields and transient responses, well-suited for the analysis of motion, and an inferotemporal stream, which contains neurons with small receptive fields, prominent color selectivity, and sustained responses, well-suited for the analysis of details of form. Visual submodalities contained within one stream appear to interact strongly, while visual submodalities in different streams appear to behave more independently (Livingstone & Hubel, 1987a,b).

This viewpoint interprets psychophysical findings in terms of the properties of cortical units in individual visual areas. The differing properties of the cortical units, in turn, follow in part from differences in their retinal inputs (Shapley & Perry, 1986): the parietal pathway may be viewed as the central continuation of the *M*-cell/magnocellular retinogeniculate pathway and the inferotemporal pathway may be viewed as the central continuation of the *P*-cell/ parvocellular pathway (Felleman & Van Essen, 1987).

One of the more striking examples of independence of visual submodalities is that of chromatic processing and motion processing. When regions are defined solely by chromatic differences, perception of motion is dramatically reduced (Ramachandran & Gregory, 1978; Cavanagh, Tyler & Favreau, 1984; Cavanagh, Boeglin & Favreau, 1985; Livingstone & Hubel, 1987b). Loss of motion perception at isoluminance is consistent with the idea that motion processing is carried out within the magnocellular pathway, while chromatic information is restricted to the parvocellular pathway. Here we ask whether regions defined solely by differences in their content of form elements support motion perception, and whether the interaction between motion and form can be simply interpreted in terms of parallel streams.

An experimental approach to this question is not straightforward. Consider a stimulus in which one region is moving with respect to another. As the two stimulus regions undergo relative motion, form elements at the boundary must change in time. This introduces an undesired cue into the stimulus: that of flicker at the boundary. Such moving, flickering regions provide strong motion percepts, in the absence
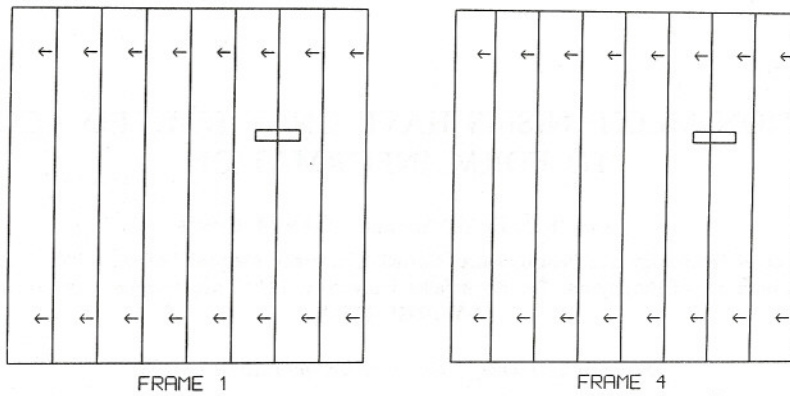
Fig. 1. A diagram of the visual stimuli used in this study. In this example, the eight vertical strips drift leftward during the stimulus presentation. Boundaries between strips are defined by combinations of flicker, granularity and form, as described in the text. The horizontal rectangle outlines the region of the stimulus detailed in Fig. 2.

of any form information (Chubb & Sperling, 1987, 1988a,b).

To eliminate cues from temporal changes in the form elements at the region boundaries, we have devised visual stimuli in which *all* form elements flicker. We used two kinds of differences between form elements: one set based on spatial frequency differences, and a second "isodipole" set which had identical spatial frequency statistics but was strikingly different in spatial structure. These stimuli thus allowed us to separate the contributions of flicker, spatial frequency content, and nonlinear feature analysis to motion perception.

METHODS

Visual stimuli

There were five kinds of visual stimuli. Each stimulus consisted of eight vertical strips which drifted uniformly left or right on successive frames (29.6 msec each). The overall layout common to all stimuli is shown in Fig. 1. All strips were identical in mean luminance and contrast. The strips and their boundaries were defined by combinations of three cues: the presence of flicker confined to the boundaries of the strips, denoted F, differences in granularity (i.e. second-order spatial correlation statistics, or, spatial power spectra), denoted G, and differences in higher-order correlation statistics, denoted H. The five stimulus types are distinguished by *which* of these cues were present: FG, G, FH, H and F. For example, the stimulus "FG" indicates that flicker and granularity cues were available to define the vertical strips and their boundaries.

Figure 2 provides details on the construction of these five stimulus types. For each stimulus type, Fig. 2 provides a detailed view of the region within the horizontal rectangle of Fig. 1 on the first four stimulus frames. In the first stimulus (FG, Fig. 2A), four of the eight strips were randomly filled with black and white checks. The intervening strips were randomly-filled with black or white rectangles, each one check wide and four checks high. Thus, for any single frame, the boundaries between the strips could be identified as the boundaries between regions of fine (1 × 1) and coarse (1 × 4) granularity. With each new frame, the boundary between strips drifted one check left (as in the example shown) or right. As this moving boundary crossed a particular check, the color-

Fig. 2 (*Opposite*). Detailed diagrams of a region of each visual stimulus used in this study. This region contains the boundary between two of the vertical strips of the complete stimulus, as illustrated in Fig. 1. This boundary drifts to the left by one check each frame. For each stimulus type, the configuration of this region on the first four frames of the stimulus is shown. Part A diagrams stimulus FG. In the initial frame shown in part A, the left half of the region is a random coloring by individual checks, and the right half is a random coloring by 1 × 4-check rectangles. On successive frames, this boundary drifts by single-check increments to the left, and only checks near the boundary are recolored. Part B diagrams stimulus G, in which the same algorithm for coloring the regions is used, but all checks are recolored on each frame. The remaining stimuli illustrated are part C: stimulus FH; part D: stimulus H; and part E: stimulus F; see text for details of the algorithms for stimulus generation.

(A)

FRAME 1:

FRAME 2:

FRAME 3:

FRAME 4:

STIMULUS FG

(B)

FRAME 1:

FRAME 2:

FRAME 3:

FRAME 4:

STIMULUS G

(C)

FRAME 1:

FRAME 2:

FRAME 3:

FRAME 4:

STIMULUS FH

(D)

FRAME 1:

FRAME 2:

FRAME 3:

FRAME 4:

STIMULUS H

(E)

FRAME 1:

FRAME 2:

FRAME 3:

FRAME 4:

STIMULUS F

Fig. 2 (a–e)

ing rule for this check was changed to make it conform to the coloring rule of its new region. Thus, checks on the boundary between strips had a probability of 0.5 of changing luminance, and checks not on the boundary did not change luminance. Since the boundary between strips moved one check with each frame, flicker of the boundary provided a second cue to the motion of the strips. This stimulus is denoted FG, to indicate that the strips were defined by two cues: flicker and granularity.

In the second stimulus, denoted G, the flicker cue was eliminated but the granularity cue was preserved (Fig. 2B). In this stimulus, all checks were recolored each frame, but the rules for coloring were the same as for stimulus FG. Thus, on any particular frame, strips were defined by their granularity (1×1 vs 1×4-check colorings), as in stimulus FG. However, since all checks were recolored each frame, all checks had a probability of 0.5 of changing luminance. Flicker as a cue to the boundaries of strips was thus eliminated.

The third and fourth stimuli, FH and H, (Figs 2C and D) were analogous to FG and G, but the "structure" cue is changed to eliminate differences in spatial-frequency content. As in the previous stimuli, one kind of strip consisted of check-by-check random colorings. The other strips were colored so that they had a striking visual structure, but nevertheless their second-order autocorrelation (and spatial frequency spectrum) matched that of the fully random coloring. The coloring rule for these strips was that of the "even" isodipole texture (Julesz, Gilbert & Victor, 1978). In this texture, the intial row $a_{i,0}$ and the initial column $a_{0,j}$ is colored at random. Interior checks are determined by the condition that all 2×2 subregions have an *even* number of white checks. The statistics of the structured strips differ from those of the random strips first at order 4. In static presentations of a single frame of this stimulus, the boundaries between the vertical strips are readily apparent. However, since extraction of the visual features that identify the structured strips requires nonlinearities in addition to spatial-frequency filtering, we denote this cue as "H" to distinguish it from the granularity cue G above.

In stimulus FH, only the checks that were crossed as strip boundaries drifted were recolored to conform with the coloring rule of their new regions. Checks at a boundary thus had a probability of 0.5 of changing in luminance,

while other checks did not change in color. Thus, for stimulus FH, this flicker provided a second cue to the boundary of stimulus regions.

In stimulus H, all checks were recolored (according to the above rules) each frame, so that all checks had a probability of 0.5 of changing in luminance. Thus, for stimulus H, the only cue available to define the regions was the presence of the high-order correlation structure of the "even" texture in one set of regions but not in the other set.

The fifth stimulus, F, contained only flicker as a cue (Fig. 2E). All regions were random colorings. On successive frames, the checks at the strip boundaries were randomly recolored, and thus had a probability of 0.5 of changing in luminance. Checks not on the boundaries were not recolored. Thus, flicker confined to strip boundaries was the only clue to motion in stimulus F. (A stimulus analogous to G or H in which both regions were random and completely recolored on successive frames would contain no cues to define strip boundaries or motion, and was therefore not included in the study except as a software check).

The stimuli described above were realized on a Tektronix 608 CRT, whose X, Y and Z (intensity) inputs were controlled by specialized electronics (modified from Milkman, Schick, Rossetto, Ratliff, Shapley & Victor, 1980) interfaced to a DEC 11/73 computer. This instrumentation provided for a 256×256-pixel raster at a refresh rate of 270 Hz. At the viewing distance of 28.5 cm, the visual display subtended a 17.6 deg square. Each hardware pixel, which subtended 4.1 min, corresponded to a "check" in the stimuli described above. Each frame of the stimulus corresponded to 8 refreshes of the raster, or 29.6 msec. The drift velocity was thus 4.1 min/29.6 msec, or 2.3 deg/sec. The luminance of the display was 150 cd/m², and the contrast $[(I_{max} - I_{min})/(I_{max} + I_{min})]$ was 0.4.

Psychophysical methods

Observers (the two investigators and two naive subjects, ages 21–33) had visual acuity of 20/20, with correction if necessary. The observer's forced-choice task was to identify the direction of motion in a presentation of a stimulus of one of the five types described above. For each stimulus type, the independent variable was the presentation time of the visual stimulus. After a warning sound, the stimulus would appear from an equiluminous uniform background for an interval of 58–947 msec, and

then the uniform background would abruptly reappear. The initial positions of the strip boundaries were randomized by trial, so that neither the initial position nor the final position of a texture strip provided any cues to direction of motion.

Observers were given practice with feedback until performance stabilized; no feedback was given during data collection. Trials were presented in blocks of 20 of a particular stimulus type and presentation time. Four blocks (80 trials) were collected for each stimulus type and presentation time. Blocks were presented in random order. Observation sessions consisted of approx. 30 blocks.

RESULTS

Qualitative findings

Informal free-view of the five stimulus types produced immediate and strong percepts of motion in the four stimuli in which either flicker or granularity was present as a cue to motion: FG, G, FH and F. This motion percept was effortless and compelling, and prolonged viewing generated a motion after-effect. Similar free-view of the stimulus H did not produce a comparably strong percept of motion. All subjects gave correct untrained judgements as to direction of motion in stimulus H, but would typically volunteer a motion percept only after several seconds of observation. Both naive subjects, as well as five additional volunteers who did not participate in the formal psychophysics, reported this judgement was distinctly harder than the other four judgements. There was no subjective difference in difficulty between the other four stimulus types for which the sense of motion was immediate.

Quantitative findings

To quantify these findings, we asked observers to make forced-choice responses to indicate the direction of perceived motion for brief presentations of the five stimulus types. Data from all subjects were in close agreement, and the features to be discussed were common to all subjects. For each observer, the mean fraction correct for each stimulus type and presentation time is shown in Fig. 3.

The most obvious difference across stimulus types was that shorter times were required for motion detection for the stimuli which contained either flicker or granularity cues. We will denote the stimulus presentation time required for 75%-correct performance (halfway between chance and perfect performance) by $t_{1/2}$. For the stimulus H whose only cue to motion was higher-order form, $t_{1/2}$ ranged from approx. 200–300 msec, and a 1000 msec presentation time was required for performance to be perfect. For the remaining four stimulus types, which contained either flicker or granularity as cues to motion, $t_{1/2}$ was 150 msec or less, and performance was essentially perfect with presentation times of 300 msec. This was in accord with the qualitative observations that these four stimulus types produced a much more salient percept of motion.

The formal psychophysics also revealed distinctions not apparent from the qualitative observations. Among the four stimuli containing either flicker or granularity cues, $t_{1/2}$ varied systematically. The shortest values were for the stimuli FG and FH containing both flicker and a second cue to motion, and were approx. 90 msec in the pooled data. For the flicker-only stimulus F, $t_{1/2}$ was approx. 120 msec. For the granularity-only stimulus, $t_{1/2}$ was approx. 140 msec.

A simple psychophysical model

In view of the consistency of the findings across subjects, we pooled the data (by simple averaging) across subjects for further analysis. The pooled data are shown in Fig. 4.

Not surprisingly, psychophysical performance was a sigmoid function of presentation time for each stimulus type. Since sigmoid curves are described not only by their halfway point ($t_{1/2}$) but also by their slope through this halfway point, one may anticipate that further information may be gleaned from the psychophysical functions by more detailed interpretation of their shape. However, such additional interpretation requires a simple model.

We consider a simple model to interpret the shape of the psychophysical curve. We postulate that there are N detectors, each acting independently, which can signal the direction of drift. The N detectors have two states: "initial" (in which they all start at the beginning of a stimulus presentation), and "set". The detectors switch from the "initial" state to the "set" state independently at a rate r while the stimulus is presented. Once in the "set" state, a detector does not revert into the "initial" state until the next stimulus presentation. We postulate that the correct psychophysical response is given if the number of detectors in the "set" state has reached a critical number N_c at the end of the

stimulus presentation. If less than the critical number N_c of detectors has reached the "set" state, the subject guesses at random (and thus has a 50% chance of being correct).

This model is clearly a caricature of what might be a more realistic picture of the psychophysical process. A more realistic picture would include provisions for fallible detectors, uncertain but not totally random decision-making when the number of detectors approaches but

does not reach N_c, and more complex kinetics. However, this simple model suffices to interpret the shape of the psychophysical curves in terms of the degree of parallelism (N and N_c) and the kinetics of individual detectors, r. The main difference between this model and more standard models in visual psychophysics is that the independent variable is time, rather than a stimulus parameter such as contrast or luminance (Johnson, 1980; Pelli, 1985). Thus, the
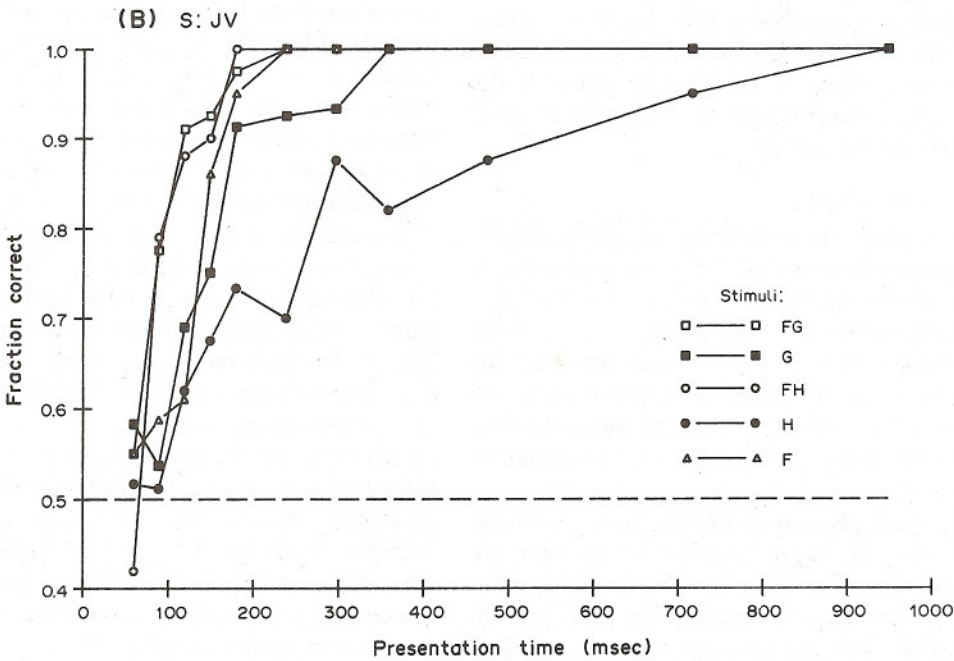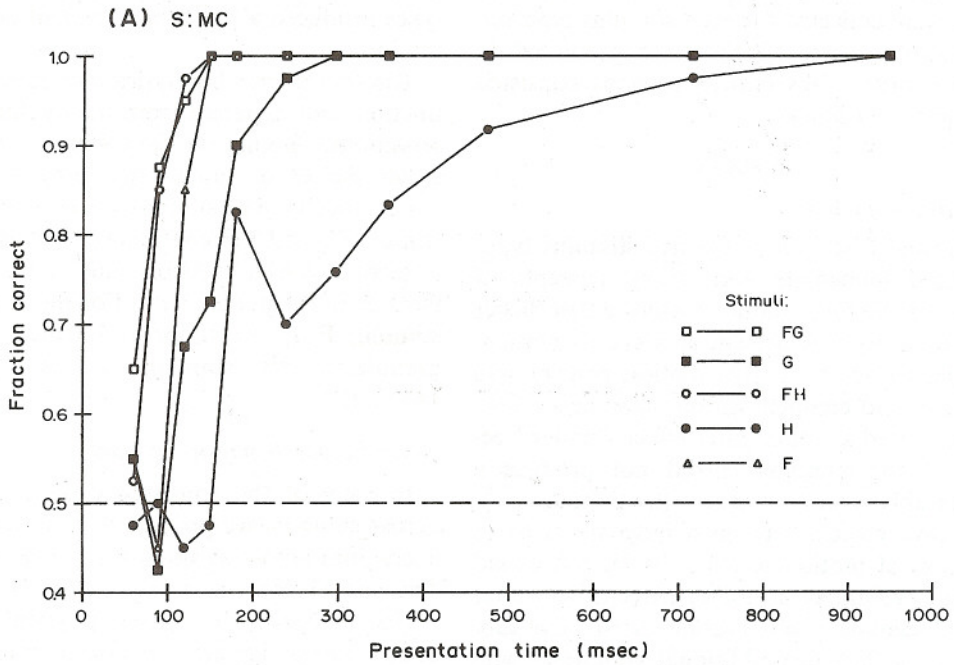

Fig. 3 (A, B)

Fig. 3. Fraction of correct judgements of direction of drift for all five stimulus types, over a range of stimulus presentation times. Observers in parts C and D are naive.

stochastic aspects of the subject's behavior are interpreted in terms of the kinetics of each detector, rather than in terms of the signal-to-noise of the sensory signal.

We now proceed to analyze this model. The crucial internal quantity is the number of detec-

tors in the "set" state at a time t after the onset of the stimulus. We denote the probability that exactly j detectors are in the "set" state at the t by $p_j(t)$. Since each detector converts from the "initial" state into the "set" state at the rate r, it follows that the probability that *all* detectors

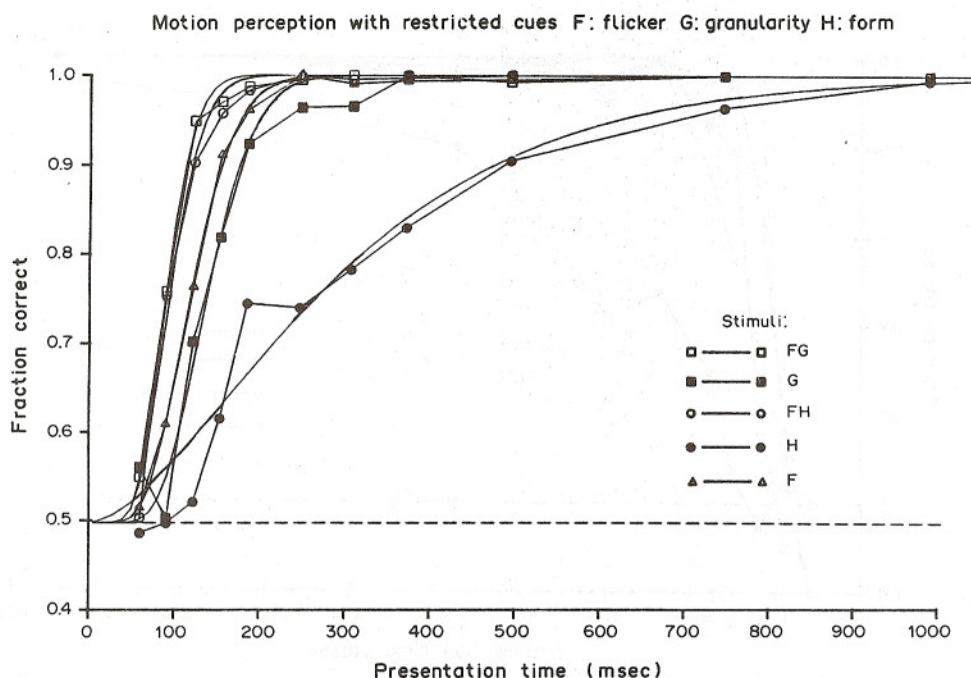Motion perception with restricted cues F: flicker G: granularity H: form

Fig. 4. Comparison of pooled data across four subjects (from Fig. 3) with the model. The smooth curves
are derived from the psychophysical function (5) and the parameters of Table 1.

are in the "initial" state at time t evolves
according to:

$$\frac{dp_0(t)}{dt} = -rNp_0(t). \qquad (1)$$

The equation for the probability $p_j(t)$ that ex-
actly j detectors are in the "set" state at time t
is slightly more complex. A condition in which
exactly j detectors are in the "set" state arises
from a condition in which $j - 1$ detectors are in
the "set" state when one of the remaining
$N - j + 1$ detectors convert. Once formed, this
condition is eliminated when an additional
detector (one of the now-remaining $N - j$
detectors) convert. Thus,

$$\frac{dp_j(t)}{\cdot dt} = -r(N - j)p_j(t)$$

$$+ r(N - j + 1)p_{j-1}(t), \quad j \geqslant 1. \qquad (2)$$

The differential equations (1) and (2) are subject
to the initial conditions $p_0(0) = 1$ and $p_j(0) = 0$
($j \geqslant 1$). They are solved by:

$$p_j(t) = \binom{N}{j} e^{-Nrt}(e^{rt} - 1)^j. \qquad (3)$$

An important limiting case is the situation in
which only a small fraction of the detectors have
been converted into the "set" state ($jrt \ll N$).

This is the Poisson limit, and the familiar
Poisson statistics are recovered:

$$p_j(t) \simeq \frac{1}{j!}(Nrt)^j e^{-Nrt}. \qquad (4)$$

According to the model, the fraction correct
$f(t)$ for a stimulus presentation of duration t is
determined by the probability that at least N_c
detectors have converted to the "set" state.
If this condition is not met, performance is
random. Thus, the psychophysical function
predicted by this model is:

$$f(t) = \sum_{j=N_c}^{\infty} p_j(t) + \frac{1}{2}\left[1 - \sum_{j=N_c}^{\infty} p_j(t)\right]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{j=N_c}^{\infty} p_j(t). \qquad (5)$$

Analysis of the model

Before we attempt to compare the model with
data, we examine how the model parameters
influence the shape of the psychophysical func-
tion $f(t)$. Two quantities are of particular inter-
est: $t_{1/2}$, the presentation time required for
75%-correct performance, and b, the slope of
the psychophysical curve at $t_{1/2}$. According to
the model, these quantities are determined by:

$$f(t_{1/2}) = 3/4 \quad \text{and} \quad b = \frac{df}{dt}(t_{1/2}). \qquad (6)$$

The only dimensional parameter of the model is r (the rate of conversion of an individual detector). This provides an overall timescale for the model, but does not otherwise influence the shape of the psychophysical curve. In dimensionless terms, the shape of the psychophysical curve is determined by the parameters N (the number of detectors) and N_c (the number of detectors required for a correct psychophysical response). We first analyze the situation $N_c = 1$, in which only a single detector is required for a correct psychophysical response. From equations (3) and (5), we find:

$$f(t) = 1 - \tfrac{1}{2}e^{-Nrt};\qquad(7)$$

and therefore that

$$t_{1/2} = \frac{\log 2}{Nr}\quad\text{and}\quad b = \frac{1}{t_{1/2}}\frac{\log 2}{4}.\qquad(8)$$

Note that model behavior does not depend separately on N or r, but only on their product Nr.

Separate dependence on N_c, N and r arises only when more than one detector must convert to the "set" state. While exact explicit formulae for $t_{1/2}$ and b cannot be obtained for these cases, approximate formulae are readily obtained in the most interesting regime: N is large, and the critical number of detectors N_c is neither very close to 1 nor to N. We call this the Gaussian regime, because the expression (3) for $p_j(t)$ is closely approximated by a Gaussian distribution of equal mean and variance. The mean $m(t)$ and variance $v(t)$ of the number of detectors in the "set" state may be calculated from equation (3):

$$m(t) = \sum_{j=0}^{\infty} jp_j(t) = N(1 - e^{-rt});\qquad(9)$$

and

$$v(t) = \sum_{j=0}^{\infty}[j - m(t)]^2 p_j(t)$$
$$= N(e^{-rt} - e^{-2rt}).\qquad(10)$$

In the Gaussian regime, we may approximate the sum in equation (5) by an equivalent integral and replace $p_j(t)$ by a Gaussian of mean $m(t)$ and variance $v(t)$. This leads to:

$$f(t) \simeq \frac{1}{2} + \frac{1}{2}\int_{N_c}^{\infty} [2\pi v(t)]^{-1/2}$$
$$\times e^{-[j - m(t)]^2/2v(t)}\,dj.\qquad(11)$$

It follows that:

$$t_{1/2} \simeq -\frac{1}{r}\log(1 - N_c/N)$$

and

$$b \simeq \frac{Nr}{2}\left(\frac{1 - N_c/N}{2\pi N_c}\right)^{1/2}.\qquad(12)$$

These expressions take a simple form if in addition it is postulated that the critical fraction of converted detectors, N_c/N, is small:

$$t_{1/2} \simeq \frac{N_c}{Nr}\quad\text{and}\quad b \simeq \frac{1}{t_{1/2}}\left(\frac{N_c}{8\pi}\right)^{1/2}.\qquad(13)$$

Although the expressions (13) are approximations which were derived from the assumptions that N is large, N_c/N is small, and that N_c is neither near 1 nor N, they remain reasonable approximations even when these conditions are not met. Even in the extreme case of $N_c = N = 1$, the approximations (13) deviate from the exact expressions (8) by only 30%. Thus, equations (13) provide an intuitive understanding of the model over a wide range of parameter values, which may be summarized as follows: performance is midway from random to perfect at the time $t_{1/2}$ which is proportional to the number N_c of detectors necessary for a decision, and inversely proportional to the population rate Nr of detector conversion. The rapidity with which psychophysical performance ascends from chance to ideal is proportional to the square root of the number of detectors required for a decision, and thus provides a crude estimate of the degree of parallelism of the process.

If N_c is comparable to the total number of detectors N, the predicted psychophysical curve is an accelerating function of presentation time, until nearly perfect performance is obtained. This concave-up shape reflects the need for nearly all detectors to convert to the "set" state prior to a decision. Our psychophysical data (Figs 3 and 4) lie on a sigmoid curve, which implies that N_c is substantially less than N. Under these conditions, N only influences the shape of the psychophysical curve to the extent that equations (12) depart from the Poisson limit (equations 13). This departure is much less than the uncertainty in the pooled psychophysical data. Thus, we reduce the number of parameters to be determined from three (N, N_c and r) to two (N_c and Nr). The new parameter $R = Nr$ is the population rate of conversion of detectors into the "set" state.

We sum up the analysis of the model. For the relevant range of parameter values, $t_{1/2}$, the predicted time required for 75%-correct performance and b, the slope of the psychophysical curve at $t_{1/2}$, depend on two quantities: N_c, the critical number of detectors required for a decision, and the parameter combination $R = Nr$, the population rate of detector conversion (equations 13).

Comparison of model with data

In view of the above analysis, we compared the measured psychophysical curves to the model (5), with the simplified expression (4) substituted for $p_j(t)$. Values of N_c and $R = Nr$ that provided the least-squares best fit to the pooled data are presented in Table 1, along with estimated values of $t_{1/2}$ and b calculated from equation (13). To reduce computations, only even values of N_c were examined for $N_c > 4$. As seen in Fig. 4, the model provides a reasonable fit to the experimental data.

We do not intend to imply that our model provides the best possible fit to the data, or that other model functions (e.g. the Weibull, 1951, function) are excluded. Rather, the purpose of the modelling is to extract as much information as possible from the shapes of the psychophysical curves. The reasonable fit of the model curves to the data thus provide the justification to consider the model parameters as a concise summary of the psychophysical data.

The values of $t_{1/2}$ presented in Table 1 describe the time at which performance has risen halfway from random to perfect. These vary systematically in the manner expected from the responses of the individual subjects (Fig. 3). The stimuli sort themselves into four sets which may be ranked in order of increasing $t_{1/2}$: {FG, FH}, {F}, {G}, {H}.

The model parameters of Table 1 reveal a second difference between the psychophysical curves. Responses to four of the stimuli (FG, FH, F and G) are well-described by psycho-

Table 1. Parameters of the model psychophysical function (5) obtained from least-squares best fits to the experimental data (Fig. 4). The Poisson regime ($N_c \ll N$) is assumed. For details, see text

Stimulus type	$R (= Nr)$ (sec^{-1})	N_c	$t_{1/2}$ (msec)	b (sec^{-1})
FG	156.1	14	89.6	16.7
G	87.6	12	137.0	10.1
FH	127.2	12	94.3	14.7
H	6.6	2	317.4	1.7
F	84.2	10	118.8	10.6

physical functions with N_c in the range 10–14. Responses to the stimulus H are only fit well with a smaller value of N_c, and are fit best with $N_c = 2$. This corresponds to a more gentle transition between chance and ideal behavior, even corrected for the more sluggish dynamics (lower R) for the stimulus H. In terms of the model, this corresponds to a smaller number of detectors responsible for the psychophysical judgement. Because of the insensitivity of the model to total number of detectors N, we cannot be sure whether the more gradual slope of the psychophysical curve also indicates a smaller total number of detectors available for the discrimination.

DISCUSSION

Summary of results

Our main findings may be summarized as follows. Moving regions whose boundaries are demarcated by flicker or differences in granularity (stimuli FG, G, FH or F) elicit strong motion percepts and reliable direction-of-motion judgements in presentations of under 150 msec. However, moving regions whose boundaries are demarcated by differences in visual features do *not* elicit strong motion percepts when flicker and spatial frequency cues are removed (stimulus H), and do *not* support reliable direction-of-motion judgements in presentations of 300 msec or less. Although the feature cue by itself provides only a very weak percept of motion, addition of this cue to a flicker-only stimulus enhances the perception of motion for stimulus durations as short as 90 msec.

Before discussing these findings in relation to general concepts of visual processing and motion detection, we first provide evidence that the lack of a strong motion percept in the stimulus H is not an artifact of the means used to eliminate the flicker cue. As pointed out in the Introduction, any stimulus in which the number of visual features in a region of space changes with time must have a component of flicker. Thus, to remove the possibility that localized flicker is available as a cue to motion, one approach is to ensure that *all* regions of the stimulus have the same likelihood of flickering. However, one may wonder that this diffuse spatiotemporal flicker will also eliminate the ability to perceive the visual features themselves. Several lines of evidence indicate that this is not the case. (i) All subjects performed at the 100% level for discrimination of the even texture from

a random texture even for single-frame (29.6 msec) masked presentations restricted to a single vertical strip of the texture. (ii) Evoked-potential studies demonstrate that the pattern-specific response to the even texture is only attenuated by about 25% when new examples of the even texture are presented for 29.6 msec, in comparison to a presentation duration of 237 msec (Victor & Conte, 1989). (iii) If diffuse spatiotemporal flicker *nonspecifically* eliminated the ability to perceive form, then it should have comparably reduced the percept of motion in stimulus G as well. However, there was no subjective reduction in the ability to see motion in this stimulus, and quantitative performance for this stimulus was only slightly worse than for the stimulus FG (Table 1). Thus, although a rather complex maneuver was necessary to eliminate localized flicker as a motion cue in a stimulus which had moving collections of local features, it is unlikely that this maneuver artifactually eliminated the ability to perceive motion. Rather, we conclude that the local features of stimulus H do not suffice for a strong motion percept. This is particularly interesting in view of the sufficiency of other local features (granularity, as in stimulus G) to drive motion perception strongly.

Motion perception in drift-balanced stimuli

The visual stimuli used in these studies are all "drift-balanced" motion stimuli (Chubb & Sperling, 1987, 1988a,b). As such, they all have the property that the basic Reichardt (1961) cross-correlation detector will not signal motion. Chubb and Sperling propose to augment the basic Reichardt detector with a nonlinear preprocessor. If (as they propose) this preprocessor were to include an element sensitive to temporal change in luminance regardless of signature, then perception of motion would be expected in all stimuli which contained localized flicker as a cue (F, FG and FH). This is indeed observed. In addition, strong motion is perceived for the stimulus G, in which localized flicker is not a cue. In this stimulus, the spatial frequency content of the regions with the 1×4 granules is *different* from the spatial frequency content of the regions with the 1×1 checks. If the preprocessor of Chubb and Sperling includes linear spatial filtering as well as rectification, strong motion perception is again predicted. However, none of these models predict motion perception for stimulus H.

Processes underlying motion detection

The ability to perceive motion in the stimuli F, FG, FH and G may be interpreted as minimal requirements for the processes which precede a Reichardt-like motion detector. These considerations apply not only for the basic crosscorrelation model of Reichardt (1961), but also for later variations on this theme (van Santen & Sperling, 1984; Adelson & Bergen, 1985).

In an analogous manner, failure to perceive strong motion in the stimulus H may be interpreted as an *upper* limit for the complexity of the processes which precede a crosscorrelator. While the precise computations that extract structure from the even texture are uncertain, they require highly nonlinear interactions among four regions of visual space (Julesz, Gilbert & Victor 1978; Victor & Zemon, 1985; Victor & Conte, 1989). The psychophysical data presented here demonstrate that although highly nonlinear interactions are critical to texture discrimination, they do not have direct input into motion detection.

Nature of the motion percept in stimulus H

Although the perception of motion from higher-order form cues alone is very weak, it is not entirely absent. Subjectively, the sense of motion builds slowly with time, simultaneous with the conscious recognition that the location of the strips containing the "even" texture change over time. Motion appears coarse, slow, and discontinuous. We speculate that the residual motion percept driven by this stimulus corresponds to a higher-level, long-range process, while the motion percept driven by the other cues is a rapid, low-level, short-range process (Braddick, 1974, 1980; Anstis, 1980). Additional support for this distinction is that stimulus H generates no motion after-effect (Anstis, 1980), while the other stimuli do. This suggests that the stimuli with flicker or granularity cues drive opponent motion mechanisms, but the stimulus H fails to do so. Secondly, the difference in the shape of the psychophysical curves (parameter N_c, Table 1) suggests that the process underlying motion perception for the flicker and granularity stimuli rely on many detectors in parallel, while the percept driven by the stimulus H relies on only a small number of detectors.

Interaction of form cues and the flicker cue

Let us examine the stimulus F in more detail (Fig. 2E). The initial frame is a random coloring of the field. The second frame differs in that one column of checks is randomly recolored. Thus, in a two-frame presentation, the borders between the strips have been demarcated by flicker, but the direction of drift is as yet undefined. Upon the transition from the second frame to the third frame, an adjacent column of checks is randomly recolored. The first cue to direction of motion is therefore present only after three frames.

Now consider the stimulus FG, which has in addition the granularity cue G. Direction of motion is cued after two frames, because the changed column of checks is associated with *one* of the regions, on the basis of its granularity. Indeed, the visual system is able to make full use of this additional information. The time $t_{1/2}$ is reduced from 118.8 to 89.6 msec, essentially one frame time.

A more surprising result is that essentially equal use can be made of the form cue H, as judged by the reduction in $t_{1/2}$ from 118.8 msec for stimulus F to 94.3 msec for stimulus FH. The pattern structure of the stimulus H is extracted only after cortical analysis of correlations among individual geniculate afferents (Victor, 1986). By itself, this cue does not drive rapid perception of motion. Nevertheless, it is available to resolve an ambiguous motion percept driven by other cues. This has implications for general concepts of visual processing, as discussed below.

Implications for the two-level concept of motion processing

Braddick (1980) discusses a two-level model of apparent motion phenomena. In this model, an initial stage of automatically-functioning elements generates a low-level motion signal. These signals provide cues (correspondence, contours) which are available to a more interpretive higher-level process. To a first approximation, the stimulus H may be viewed as tapping only the higher-level process: the stimulus was designed so that cross-correlations provide *no* cue to boundary or direction of motion, motion perception is slow in onset, and the stimulus generates no motion after-effect. However, the interaction of its motion cue (H) and the flicker cue (F) suggests that lower-level

motion computations may be aided by the "later" higher-level computations.

Implications for concepts of parallel pathways

A dissociation between motion processing and chromatic processing is well-documented (Cavanagh et al., 1984, 1985; Livingstone & Hubel, 1987b). In analogy with the reduction but not elimination of motion perception in stimulus H, motion perception is dramatically reduced under appropriate isoluminant conditions. Nevertheless, chromatic cues are used in motion processing (Krauskopf, Farell & Movshon, 1989). In particular, chromatic cues may remove the ambiguity in an otherwise-ambiguous apparent-motion stimulus (Papathomas, Gorea & Julesz, 1989).

Similarly, it is difficult to interpret the present data simply as evidence that form (even if achromatic) and motion are processed independently. This is because the addition of the form cue to a flicker-only stimulus significantly improves perception of motion. Addition of the form cue shortens $t_{1/2}$, the time for the half-maximal performance, from 118.8 msec (stimulus F) to 94.3 msec (stimulus FH) (Table 1), even though the form cue by itself produces performance at chance level until presentation time reaches 150 msec. For these interactions of visual submodalities, a view more complex than that of independent parallel processing is required.

REFERENCES

Adelson, E. H. & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A2*, 284–299.
Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society, London, B 290*, 153–168.
Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14*, 519–529.
Braddick, O. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society, London, B 290*, 137–151.
Cavanagh, P., Tyler, C. W. & Favreau, O. E. (1984). Perceived velocity of moving chromatic gratings. *Journal of the Optical Society of America, A1*, 893–899.
Cavanagh, P., Boeglin, J. & Favreau, O. E. (1985). Perception of motion in equiluminous kinematograms. *Perception, 14*, 151–162.

Chubb, C. & Sperling, G. (1987). Drift-balanced random stimuli: A general basis for studying non-Fourier motion mechanisms. *Investigative Ophthalmology and Visual Science, 28* (Suppl.), 233.

Chubb, C. & Sperling, G. (1988a). Processing stages in non-Fourier motion perception. *Investigative Ophthalmology and Visual Science, 29* (Suppl.), 266.

Chubb, C. & Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America, A5,* 1986–2006.

Felleman, D. J. & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology, 57,* 889–920.

Hubel, D. H. & Livingstone, M. S. (1987). Segregation of form, color and stereopsis in primate area 18. *Journal of Neuroscience, 7,* 3378–3415.

Johnson, K. O. (1980). Sensory discrimination: Decision process. *Journal of Neurophysiology, 43,* 1771–1792.

Julesz, B., Gilbert, E. & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics, 31,* 137–140.

Krauskopf, J., Farell, B. & Movshon, J. A. (1989). Phenomenal coherence of moving chromatic gratings. *Investigative Ophthalmology and Visual Science, 30* (Suppl.), 389.

Livingstone, M. S. & Hubel, D. H. (1987a). Connections between layer 4B of area 17 and thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *Journal of Neuroscience, 7,* 3371–3377.

Livingstone, M. S. & Hubel, D. H. (1987b). Psychophysical evidence for separate channels for the perception of form, color, movement and depth. *Journal of Neuroscience, 7,* 3416–3468.

Milkman, N., Schick, G., Rossetto, M., Ratliff, F., Shapley, R. & Victor, J. D. (1980). A two-dimensional computer-controlled visual stimulator. *Behaviour Research, Methods and Instrumentation, 12,* 283–292.

Papathomas, T. V., Gorea, A. & Julesz, B. (1989). The strength of color and luminance in eliciting motion perception. *Investigative Ophthalmology and Visual Science, 30,* (Suppl.), 388.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, A2,* 1508–1532.

Ramachandran, V. S. & Gregory, R. L. (1978). Does colour provide an input to human motion perception? *Nature, London, 275,* 55–56.

Reichardt, W. (1961). Autocorrelation, a principie for the evaluation of sensory information by the central nervous system. In Rosenbluth, W. A. (Ed.), *Sensory communication.* New York: Wiley.

Shapley, R. & Perry, V. H. (1986). Cat and monkey retinal ganglion cells and their functional roles. *Trends in Neuroscience, 9,* 229–235.

Shipp, S. D. & Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature, London, 315,* 322–324.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194–214.

Van Essen, D. C. & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neuroscience, 6,* 370–375.

Van Essen, D. C., Newsome, W. T., Maunsell, J. H. R. & Bixby, J. L. (1986). The projections from striate cortex (V1) to visual areas V2 and V3 in the macaque monkey: Asymmetries, areal boundaries, and patchy connections. *Journal of Comparative Neurology, 244,* 451–480.

Van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America, A1,* 451–473.

Victor, J. D. (1986). Isolation of components due to intracortical processing in the visual evoked potential. *Proceedings of the National Academy of Science, U.S.A., 83,* 7984–7988.

Victor, J. D. & Conte, M. M. (1989). Cortical interactions in texture processing: Scale and dynamics. *Visual Neuroscience, 2,* 297–313.

Victor, J. D. & Zemon, V. (1985). The human visual evoked potential: Analysis of components due to elementary and complex aspects of form. *Vision Research, 25,* 1829–1942.

Weibull, W. (1951), A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18,* 292–297.

Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research, 53,* 422–427.

Zeki, S. M. (1977). Colour coding in the superior temporal sulcus of rhesus monkey visual cortex. *Proceedings of the Royal Society, London, B 197,* 195–223.