Vision Res. Vol. 33, No. 14, pp. 2021–2030, 1993
Printed in Great Britain

0042-6989/93 $6.00 + 0.00
Pergamon Press Ltd

Nonlinear Combination Rules and the Perception of Visual Motion Transparency

JEFFREY B. MULLIGAN*

Received 7 February 1992; in revised form 15 October 1992

Experiments were performed to elucidate the decomposition performed by the human visual system in the segregation of complex motion stimuli into distinct moving surfaces. Subjects were presented with achromatic patterns consisting of four types of elements, generated from two random binary luminance patterns (random-dot checkerboards). The luminance of each of the four region classes was under program control. Animated sequences of such images were produced by displacing each of the two generating patterns in opposite directions on a frame by frame basis. These displays evoke a wide variety of percepts, depending on the programmed luminance values, including motion in a single direction, simultaneous motions of transparent sheets in opposite directions, dynamic noise with no directional component, or any combination of the above percepts. A theory is presented which relates the strengths of these percepts to the amplitudes of the components in the perceptual decomposition. The experiments described measured thresholds for seeing noise or "twinkling" in addition to the multiple motions, with the goal of determining the particular signal transformations preceding motion analysis. The results are consistent with a motion extraction mechanism which operates on a linear representation of the input imagery. These results extend a similar finding due to Anstis and Mather [(1985) *Perception, 14,* 167–179] and call into question the interpretation of a recent study by Stoner, Allbright and Ramachandran [(1990) *Nature (London), 344,* 153–155].

Motion perception   Linearity   Transparency

INTRODUCTION

Human observers are sensitive to visual motion over a wide range of spatial and temporal parameters, despite the fact that animal studies indicate that individual neurons early in the visual pathway only respond over a restricted range of parameters. It can as yet only be speculated how motion information from different spatial and temporal "channels" might be integrated by decision-making processes. This process is further complicated by the fact that the visual system is not merely concerned with detecting isolated events, but must interpret a complex input stream generated by a multiplicity of objects in the environment. This paper attempts to shed some light on the details of the mechanism by which the brain is able to segment individual objects, even when they occupy the same region of visual space, as occurs in the phenomenon known as *visual transparency*.

The word "transparency" has a slightly different connotation in the study of vision than in colloquial usage. One says a pane of glass is transparent because one can see through it; the term visual transparency, on the other hand, refers to a situation where one sees *both* the glass *and* the objects behind it, which is perhaps more properly described by the term "translucency". Here the term visual transparency is used in a broad sense to describe a situation where multiple surfaces or objects are perceived at a single visual location.

It is possible to evoke a perception of transparency in static images (Tudor-Hart, 1928), and some general constraints on the luminance relations have been described (Metelli, 1970, 1974, 1985; Metelli, Da Pos & Cavedon, 1985; Beck, 1978, 1985, 1986; Beck, Prazdny & Ivry, 1984; Brill, 1984, 1986; Masin, 1984; Beck & Ivry, 1988; Adelson, 1990; Kersten, 1991). The perception of transparency can often be enhanced when differences in motion help to segregate the two surfaces. It is not strictly necessary to have two distinct surfaces to produce motion transparency; e.g. a shadow may move across an object, yet the motion of the shadow, although it is seen, is not attributed to the underlying object, nor is it attributed to the presence of a second surface. I will use the term *motion transparency* to refer to a situation where multiple distinct motions are seen at a single location; this may or may not be accompanied by the perception of multiple surfaces at different depths.

Motion transparency can arise in the natural world in a number of ways. The example given above of shadows is one of the simplest and most common examples. In this case, the stimulus can be described as the *product* of a reflectance image (i.e. the object's texture and/or coloration) and an illumination image (i.e. the spatial pattern of light and shadow). (In this paper the case of

*NASA Ames Research Center, Mail Stop 262-2, Moffett Field, CA 94035, U.S.A.

colored shadows, which is significantly more complicated, will not be considered.) This simple mathematical relationship also describes the physics of partially transparent objects when there is no diffuse reflection, such as a projected sandwich of black-and-white transparencies. I shall refer to this situation as *multiplicative transparency*.

Another example is that of specular reflection from glossy surfaces such as leaves, fur, skin, hair, or cellophane. Specular reflection from a planar surface, such as an air-water interface, produces a stimulus in which a reflected image is combined *additively* with the images of submersed objects. On irregular surfaces, such as hair, the specular reflection usually takes the form of a distorted image of the light source, often called a "highlight". In either case, the motion of the reflected image is in general different from that of the reflecting object (or the underwater objects in the case of water reflection). I shall refer to this situation as *additive transparency*.

The final case to be considered occurs when a small opaque object or collection of objects moves in front of a larger object. Although this is not transparency in the strict sense, the larger occluded object is often perceived without interference from the small occluders. Indeed, a sensor tuned to low spatial frequencies will respond to the large object and may be totally blind to the small occluders. One example is when an object is viewed from behind a finely spaced network of twigs and branches. I shall use the term *transparent occlusion* to refer to this situation. In Fig. 1, numerical examples are shown for these three distinct physical mechanisms in the special case where each of the two patterns being combined is restricted to only two levels.

Transparency has also been observed in the laboratory using synthetic stimuli which were created to study

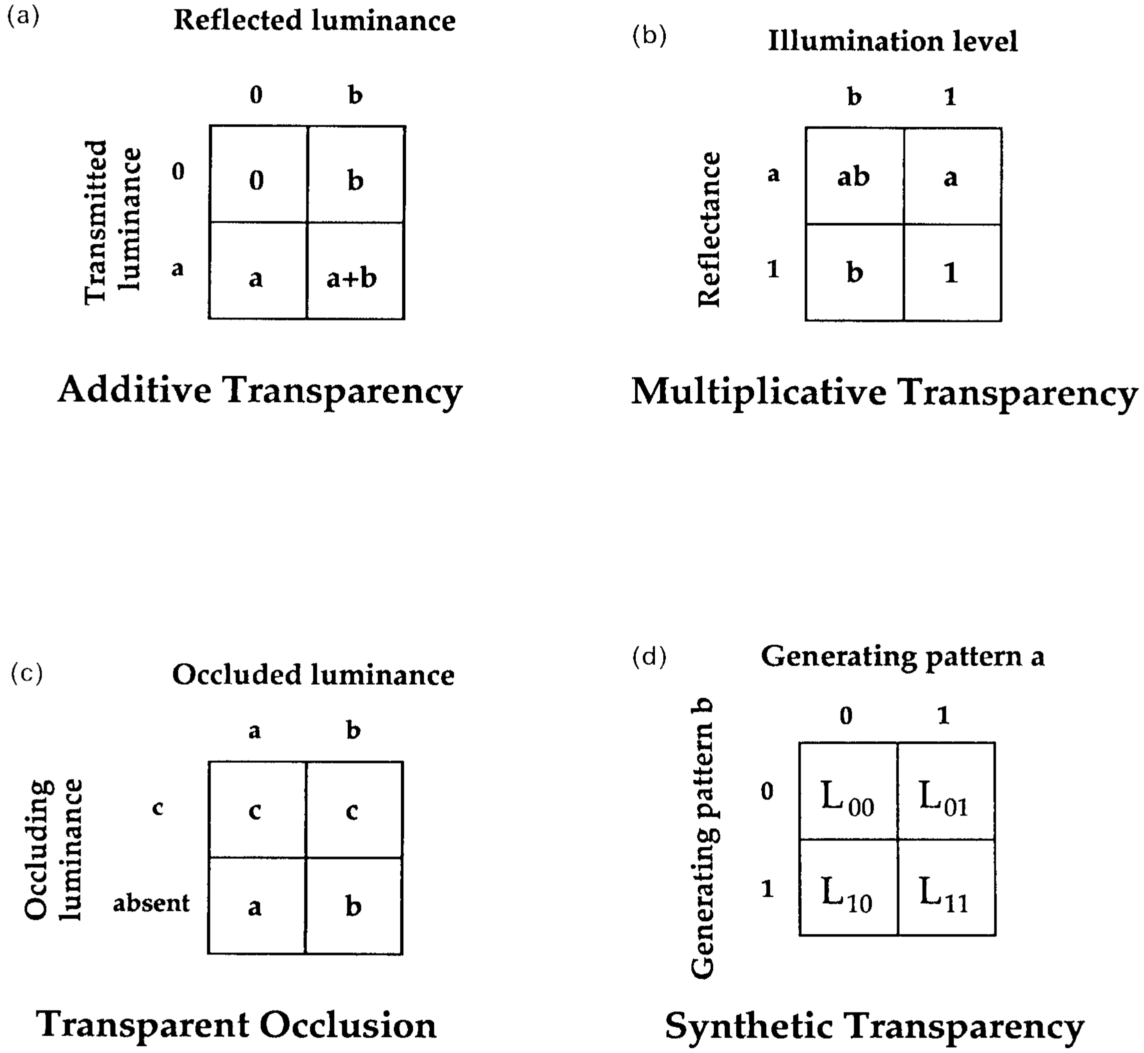

FIGURE 1. Examples of the luminances resulting from the combination of binary generating patterns using rules corresponding to a variety of physical mechanisms. (a) Additive combination of two binary patterns, each of whose elements are either black or a particular gray level. These luminances would result if the two patterns were combined by shining two projectors onto the same screen. (b) Luminances arising from the multiplicative combination of two binary patterns, such as occurs when one pattern is used to illuminate another. (c) Luminances arising from transparent occlusion of one pattern by another. (d) When synthetic stimuli are generated by computer, one is not constrained to simulate physical combination rules, and the four luminances in the matrix may be set to arbitrary values [which may or may not correspond to one of the situations diagrammed in (a)–(c)].

phenomena other than transparency. Adelson and Movshon (1982) used an additive combination of sine-wave gratings differing in orientation (a plaid) to study integration of motion signals, and observed that the stimulus might be seen either as a single "coherent" pattern, or "incoherently" as two individual components with different motions. The term "incoherent" in the plaid literature is synonymous with what is here called motion transparency. A recent study by Stoner, Allbright and Ramachandran (1990) investigated the dependence of plaid coherence upon the rule used to combine the component luminances. In previous studies, sine-wave plaids have been created simply by adding the component luminances. Stoner *et al.* used square-wave gratings, both to enhance the perception of transparency, and to simplify the simulation of a variety of physical situations, since the resulting plaid patterns could be made up from only four gray levels. (When the components have equal contrasts, the number of gray levels degenerates to three.) They found that they could influence the degree to which transparency was seen by manipulating the luminance of the intersection regions, and that transparency (i.e. component motion) was most likely to be seen when the luminances satisfied the physical constraints for multiplicative transparency or translucency (with the possibility of a diffuse reflection component).

The results of Stoner *et al.* are somewhat surprising, since the multiplicative plaids actually contain Fourier components moving in the pattern direction. The puzzle traditionally associated with plaid stimuli is how the visual system determines the "pattern direction", which is seen when the plaid is perceived as a single moving surface, since the earliest direction-selective neurons are presumed to respond only to the individual Fourier components; when a plaid is made additively there are no Fourier components oriented perpendicular to the pattern motion direction, and so cells in primary visual cortex which respond to their preferred orientation moving in the direction of the pattern velocity will not be activated by the plaid stimulus (assuming the pattern direction is sufficiently different from the component directions). It is somewhat paradoxical that adding a stimulus component which by itself would be seen moving in the pattern direction should inhibit the perception of coherent pattern motion!

A key assumption that is seldom stated is that a *linear* representation of the stimulus is maintained up until the site of motion analysis. If the plaid is subjected to a logarithmic nonlinearity prior to motion analysis, however, then the extra Fourier components moving in the pattern direction in a multiplicative plaid will be removed. Such a nonlinearity would similarly *introduce* additional Fourier components in the form of distortion products to the representation of a stimulus created additively. Since direction-selective cells in striate cortex do *not* generally respond to the pattern motion (Movshon, Adelson, Gizzi & Newsome, 1986), one might be tempted to conclude that there are no significant nonlinear distortions on the inputs to these cells (at least for the contrasts investigated), although the data have not been analyzed specifically to test this hypothesis.

The question of whether the motion system deals with a linear representation of the stimulus is also germane to the motion analysis of unambiguous transparent stimuli, such as shadows and highlights. When signals combine additively, as in the case of specular reflection, motion analysis following a bank of narrowly tuned linear spatio-temporal filters will automatically segregate signals originating from the reflected and transmitted images (assuming these component images have sufficiently different motions). Stimuli arising as the product of two components, such as shadow illumination of a texture, will contain distortion products that generate spurious motion signals, and which are inconsistent with the motion of either component. If, on the other hand, the stimuli are passed through a logarithmic transform prior to filtering and motion analysis, then the product is transformed to a sum, and the shadows may now be treated as was originally suggested that reflections might be. Such a compressive nonlinearity would now introduce distortion products into the stimulus arising as the additive combination of component patterns. Thus it would seem that it is difficult to design a system which can easily deal with both types of motion transparency which commonly occur in natural scenes. The experiments described in this paper were performed in an attempt to determine which of the two strategies described above might be employed by the human visual system.

The general approach employed in the present study was to synthesize stimuli corresponding to a variety of combination rules, including ones corresponding to the simple physical mechanisms described above. A base pattern was made by combining oppositely moving random dot textures; the luminances of the individual elements could then be set in different ways to simulate the different combination mechanisms. Within this framework, it was also possible to create a wide range of stimuli which did not correspond to any simple physical mechanism. Preliminary observations indicated that stimuli simulating additive or multiplicative transparency generally evoked a percept of two sheets of random texture sliding smoothly over one another; some synthetic stimuli, however, evoked a sensation of noise or "twinkling" in addition to the two transparent motions. This suggested the hypothesis that the motion system was performing a particular decomposition on the stimulus, and that the part of the stimulus which was leftover after the leftward and rightward components had been categorized was responsible for the perception of "twinkle".

STIMULI

Unlike plaids, the transparent stimuli used in the present study were made by combining two-dimensional components whose individual motions were *not* ambiguous; in this case the resulting stimulus has no physically plausible "coherent" interpretation. The stimuli were created by combining two generating patterns, each of which was a random binary luminance pattern (random-dot checkerboard). In these patterns, the state of each square element is chosen independently to be either 0 (black) or 1 (white) with equal probability. A typical example of one of these generating patterns is shown in Fig. 2(a). The generating patterns were computed on a 64×64 grid, as is shown in the figure.

Composite images were made by using the values of the binary generating patterns as the digits of a two-bit binary number, resulting in values between 0 and 3 (inclusive). Successive frames in the sequence were created by spatially shifting each of the generating patterns by one check-width (with wrap-around) prior to combination. One of the generating patterns was moved from left-to-right, while the other was moved in the opposite direction at an equal rate. I shall use the symbol $\mathbf{L}_{ij}$ to refer to the luminance of those regions having a value expressed by the binary digits i and j.

A useful way of describing the stimuli is as follows: let $\mathbf{i}(x, y, t)$ represent a given stimulus, and let $\mathbf{a}(x, y)$ and $\mathbf{b}(x, y)$ be the binary generating patterns used to create it. It is convenient to assume that the binary values of the generating patterns $\mathbf{a}$ and $\mathbf{b}$ represent the values ± 1, so that they represent a pure contrast signal with no d.c. component. The stimulus $\mathbf{i}$ can then be expressed in terms of the contrasts of the generating patterns:

$$\mathbf{i}(x, y, t) = \mathbf{L}_{\text{mean}}[1 + \mathbf{C}_{\text{a}}\mathbf{a}(x - t, y) + \mathbf{C}_{\text{b}}\mathbf{b}(x + t, y) + \mathbf{C}_{\text{ab}}\mathbf{a}(x - t, y)\mathbf{b}(x + t, y)]$$

where

$$\mathbf{L}_{\text{mean}} = \frac{\mathbf{L}_{11} + \mathbf{L}_{10} + \mathbf{L}_{01} + \mathbf{L}_{00}}{4}$$ the mean luminance;

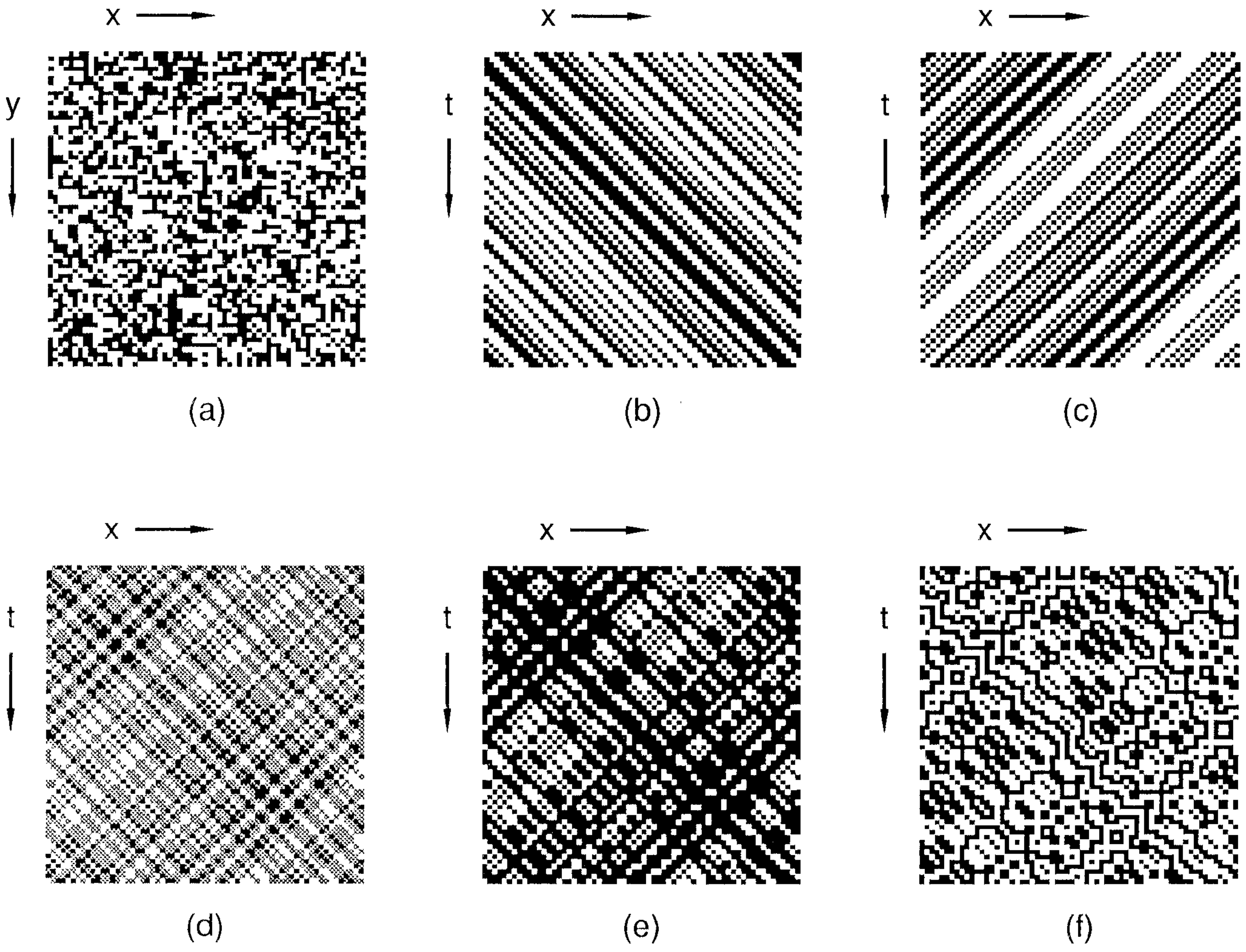

FIGURE 2. (a) A representative binary generating pattern. (b)–(f) Space–time plots of a single scan line of a single animated sequence resulting from different assignments of the pixel luminances. In (b) and (c), the luminances are set so that only a single direction of motion is visible. In (d), the two generating patterns are combined additively; both leftward and rightward motions are seen simultaneously, corresponding to the two prominent orientations visible in the figure. In (e), the luminances are set using an occlusion rule [which corresponds to taking the logical-AND of (b) and (c)]. When presented with this stimulus, subjects report seeing both directions of motion [as with the stimulus illustrated in (d)], *plus* an additional component of noise or "twinkle". When the generating patterns are combined as is shown in (f), which is the logical exclusive-OR of (b) and (c), *only* twinkle is seen. The image in (e) can be described as a linear combination of the images in (d) and (f).

$$C_a = \frac{L_{11} - L_{10} + L_{01} - L_{00}}{4L_{mean}}$$

the contrast of rightward-moving pattern;

$$C_b = \frac{L_{11} + L_{10} - L_{01} - L_{00}}{4L_{mean}}$$

the contrast of leftward-moving pattern;

$$C_{ab} = \frac{L_{11} - L_{10} - L_{01} + L_{00}}{4L_{mean}}$$

the contrast of the product pattern **ab**.

Some illustrative examples are shown in Fig. 2. Figure 2(a) shows a typical noise pattern used for the generating patterns **a** and **b**. The remainder of the panels in Fig. 2 are space time plots of a single scan line of the composite pattern, with the space axis running horizontally, and the time axis running vertically, with time increasing from top to bottom. Figure 2(b) illustrates the case $L_{00} = L_{10} = 0$, and $L_{01} = L_{11} = 1$, for which C_a, the contrast of the rightward moving pattern **a** has a value of 1, and C_b and C_{ab} are zero. The converse case where $C_b = 1$ and $C_a = 0$ is shown in Fig. 2(c). It will be remembered that in space time plots such as these, orientation corresponds to velocity; hence the strongly oriented patterns in Fig. 2(b, c) indicate the motions of the generating patterns.

Figure 2(d) shows the case $L_{11} = 1$, $L_{00} = 0$ and $L_{10} = L_{01} = \frac{1}{2}$. This corresponds to additive transparency, which would be obtained by combining the patterns in Fig. 2(b, c) with a beam splitter. In this case $C_a = C_b = \frac{1}{2}$, and $C_{ab} = 0$. Note that in the figure both orientations are clearly seen; when the animated sequence is viewed, both directions of motion are seen.

Figure 2(e) shows the case $L_{00} = L_{01} = L_{10} = 0$, $L_{11} = 1$. This corresponds to the logical AND of the two patterns and can be interpreted as the dark elements of one pattern occluding the elements of the other. In this case, $C_a = C_b = C_{ab} = 1$. As in Fig. 2(d), visual inspection of the space time plot reveals that both orientations are clearly visible, and when the animated sequence is viewed both directions of motion are seen. In this display, however, most subjects report that the pattern "twinkles" while the motion is being depicted. The central hypothesis of this paper is that this perception of twinkle is directly related to the nonlinear distortion product having contrast C_{ab}. A space time plot of one scan line of this distortion product is shown in Fig. 2(f). This pattern is obtained by setting $L_{00} = L_{11} = 0$, and $L_{01} = L_{10} = 1$. In this case, $C_a = C_b = 0$, and $C_{ab} = 1$. Note that if the values of **a** and **b** are defined to be 0 and 1 (instead of ± 1) then the "product" pattern could be computed as the logical exclusive-OR of the generating patterns. Although the pattern appears structured when compared to Fig. 2(a) (which as a space time plot would represent dynamic random noise), the conspicuous oriented features which are visible in Fig. 2(d, e) are now absent. Structures oriented at ± 45 deg can be seen, but

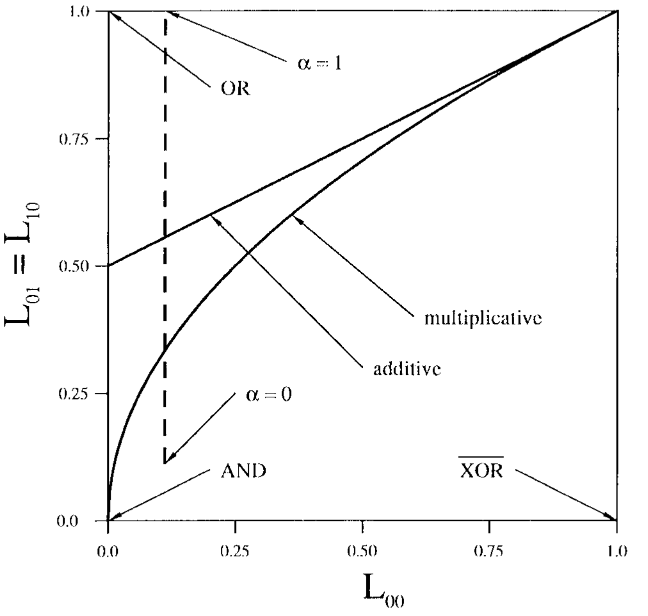

FIGURE 3. A two-dimensional slice through the four-dimensional stimulus space generated by varying the element luminances. L_{11} is assumed to have a constant value of 1. The abscissa represents the value of L_{00}, while the ordinate represents the values of L_{01} and L_{10}, which are assumed to be equal. The linear locus with a slope of $\frac{1}{2}$ corresponds to stimuli generated as additive combinations of the generating patterns; the leftmost point of this locus corresponds to the stimulus depicted in Fig. 2(d). The parabolic locus corresponds to stimuli generated as multiplicative combinations of the generating patterns; the leftmost point of this locus represents the stimulus shown in Fig. 2(e), which can also be computed as the logical AND of the generating patterns. The point at the lower right hand corner of the figure, labeled XOR, represents the stimulus which is the complement of the exclusive-OR of the generating patterns [shown in un-complemented form in Fig. 2(f)]. The vertical dashed line on the left side of the figure corresponds to the stimuli presented in the experiment. This line segment is parameterized by α, which was varied between 0 and 1. The figure illustrates that as α is varied, both the multiplicative and additive forms of transparency can be obtained.

these structures are second-order textural features. This can be seen by observing the figure with optical defocus; when the image is blurred, the oriented texture patterns become hard to see, and the figure becomes hard to distinguish from the completely random pattern in Fig. 2(a). As might be expected from these observations, when this product pattern is viewed as an animated sequence, little coherent motion is seen; in fact, informal observations suggest that for short durations (100 msec or less) sequences of this type cannot be discriminated from actual dynamic random noise.

A slice of the full four dimensional parameter space is shown in Fig. 3. In Fig. 3 it is assumed that L_{11} is held constant to a value of 1, and that L_{01} and L_{10} are equal. Each point in the plot corresponds to a particular pair of values for L_{00} and L_{01}. The pattern formed by the logical AND of the two generating patterns, which was shown in Fig. 2(e), is represented by the point in the lower left corner. The point in the lower right corner represents the complement of the exclusive-OR pattern from Fig. 2(f). The line with the slope of $\frac{1}{2}$ is the locus of points which correspond to purely additive combinations of the generating patterns; different points on the line have different mean luminances and different contrast values. The parabolic curve corresponds to purely multiplicative combinations.

It is possible to label each point in Fig. 3 with the percept evoked by the corresponding stimulus. Stimuli represented by the multiplicative and additive loci are generally perceived to consist of rightward and leftward motions, with no superimposed flicker or twinkle. Stimuli represented at the upper and lower left corners of the plot (AND and OR) also evoke the perception of both leftward and rightward motion, but in addition are generally perceived to twinkle. The XOR stimulus represented by the point in the lower right hand corner of Fig. 3 does not evoke any coherent motion percept, but simply appears to flicker.

These observations suggested the following hypotheses: first, that the perception of motion to the right occurs when the contrast $\mathbf{C}_a$ exceeds a certain threshold, the precise value of which is subject to masking effects due to $\mathbf{C}_b$ and $\mathbf{C}_{ab}$ (with an identical dependence of the perception of leftward motion upon the value of $\mathbf{C}_b$). Secondly, that the perception of noise or twinkle occurs when the value of $\mathbf{C}_{ab}$ exceeds another threshold, which again is subject to masking by $\mathbf{C}_a$ and $\mathbf{C}_b$. Lastly, the values $\mathbf{L}_{ij}$ which are used to compute these contrasts for the purposes of predicting thresholds are not the raw screen luminances, but rather *the luminances transformed by any nonlinear processes which precede motion analysis in the visual system*. The rationale for the final hypothesis is that if the visual system were optimized for perceiving multiplicative transparency (as occurs with shadows), then it might be useful to precede motion analysis by a logarithmic nonlinearity; after such a nonlinearity, independent motions of shadows and shaded objects could be recovered by linear spatio-temporal filters, such as those proposed by Watson and Ahumada (1985), and Adelson and Bergen (1985). The purpose of the experiments described in this paper was to gather evidence concerning such an early nonlinearity by determining which of the stimuli represented in Fig. 3 evoked the minimal amount of the "twinkle" percept. In the case of no early nonlinearity, the twinkle free zone would be expected to straddle the linear "additive" locus of Fig. 3. In the presence of logarithmic compression, on the other hand, it would be expected to follow the parabolic "multiplicative" locus.

METHODS

The stimuli were generated using an Adage RDS3000 digital raster graphics system under the control of a PDP11/73 computer. A single sequence made from a pair of generating patterns was used for the entire experiment. The index of the color of each region (0–3) was stored in the frame buffer memory; the luminance associated with each color was controlled by writing the corresponding entry in a hardware color lookup table (LUT). Each frame in the sequence was stored in a 64×64 array of pixels in the frame buffer memory. A hardware zoom feature was exploited to allow this small amount of memory to fill the entire display screen. This allowed all 64 frames in the cycle to be stored in the frame buffer simultaneously; the frames could then be cycled by changing the settings of hardware pan and scroll registers. The frames were presented at a rate of 60 Hz with no interlace.

The stimuli were displayed on a 19 in. video monitor (Mitsubishi, model C-3919N/C). Calibration was performed using a photodiode (United Detector Technology, detector head model 248, optometer model 61), equipped with a photometric correction filter and a lens which imaged the screen on the photodiode in order to measure luminance. This was done with full-screen uniform fields at each setting of the hardware lookup table. The resulting data were transformed to log log coordinates, where a piecewise linear fit was performed to generate a smooth function relating display luminance to input setting. This function was then inverted to generate a software table used to convert desired luminance values to hardware settings. This mapping was done at a fairly low software level, so the experimenter was able to deal exclusively in luminance units without concern for the mechanics of gamma correction.

The viewing distance was 3 m, from which the small square elements making up the patters subtended 5 min arc, making the total extent of the display slightly greater than 5 deg. Since the patterns moved by one element each frame transition, and the frame rate was 60 Hz (noninterlaced), the resulting drift velocities were ± 5 deg/sec. The stimuli were presented for a duration of 1 sec. Subjects were instructed to look directly at the stimuli; no attempt was made to insure that the subjects maintained steady fixation, so it is likely that during the course of the trials that some pursuit eye movements were made in response to one or both of the motions. No attempt was made to fix or stabilize the head.

The only parameter which was varied on a trial-by-trial basis was the assignment of luminances to each of the four pixel types (accomplished by reprogramming the video LUT). Each stimulus could then be described by four numbers, representing the luminances of each of the four species of pixel.

In the experiment, $\mathbf{L}_{11}$ was held fixed at a value of 100 cd/m^2, and $\mathbf{L}_{00}$ was held fixed at a value equal to one-ninth of this or 11 cd/m^2. The constraint that $\mathbf{L}_{10} = \mathbf{L}_{01}$ was also imposed; this luminance was varied from trial to trial. Note that when $\mathbf{L}_{10}$ and $\mathbf{L}_{01}$ equal the arithmetic mean of $\mathbf{L}_{11}$ and $\mathbf{L}_{00}$, then the case of additive transparency is obtained, whereas when they equal the geometric mean we obtain the case of multiplicative transparency. Trial values for $\mathbf{L}_{10}$ and $\mathbf{L}_{01}$ were obtained using the formula

$$\mathbf{L}_{01} = \mathbf{L}_{10} = \alpha \mathbf{L}_{11} + (1 - \alpha)\mathbf{L}_{00},$$

where the parameter α was varied between 0 and 1, sampled at intervals of 0.025. Additive transparency is obtained when $\alpha = \frac{1}{2}$, regardless of the values of $\mathbf{L}_{00}$ and $\mathbf{L}_{11}$. The value of α corresponding to the geometric mean, on the other hand, *does* depend on $\mathbf{L}_{00}$ and $\mathbf{L}_{11}$; for the present case, where $\mathbf{L}_{00} = \mathbf{L}_{11}/9$, the geometric mean has a value of $\mathbf{L}_{11}/3$, which corresponds to $\alpha = 0.25$. The stimuli generated under these conditions are represented in Fig. 3 by the vertical dashed line at the left side of the figure.

Of course, this one-dimensional slice through the four-dimensional parameter space of luminance settings

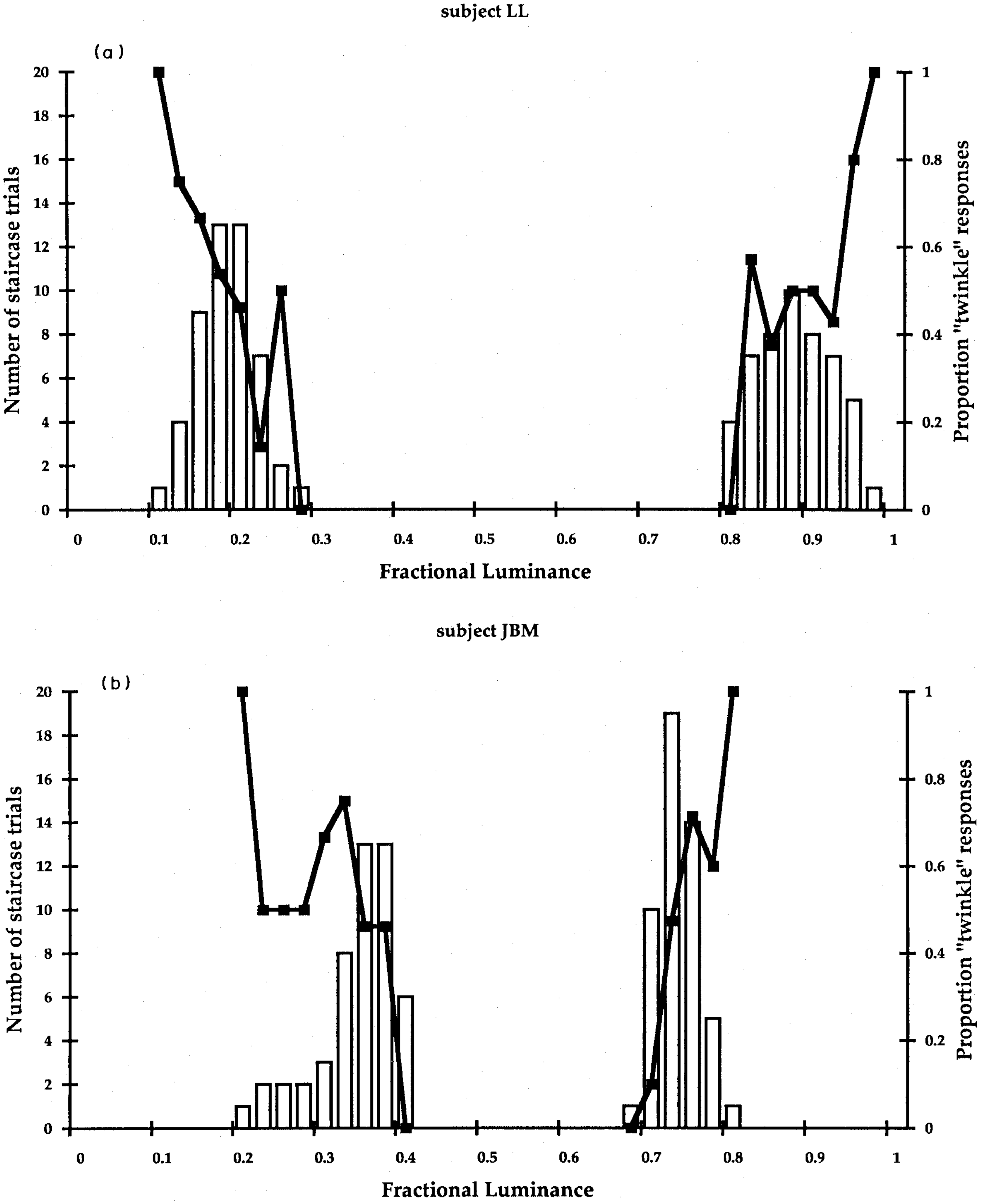

FIGURE 4. Raw data from a typical block of 100 trials for subjects JBM and LL. The abscissa is the fractional luminance parameter α. The vertical bars plot the number of staircase trials at each sampled value, indicated by the scale at the left of the figure. The squares connected by heavy lines show the proportion of trials for which the subjects reported seeing "twinkle" in addition to the two motions.

represents only a small fraction of the possible stimuli. Preliminary investigations revealed however, that this subset encompassed the perceptual categories of interest: for intermediate values of α, smooth motion transparency with no "twinkle" was seen, whereas subjects generally agreed that the stimulus seemed to twinkle at the extremes $\alpha = 0$ (dark occluders) and $\alpha = 1$ (light occluders).

On each trial subjects were asked to respond whether or not they saw twinkling in addition to motion. The value of the parameter α on each trial was controlled by a staircase procedure. The values were sampled from a discrete set which ranged between 0 and 1 in uniform increments of 0.025. Two staircases were run, one of which was started at $\alpha = 0.25$, and the other at $\alpha = 0.75$. The staircase which started at $\alpha = 0.25$ was designed to increase the variable luminance if the subject reported twinkle, and to decrease it otherwise. The other staircase which was started at $\alpha = 0.75$ responded with the opposite type of feedback. Thus the two staircases traced out two complementary limbs of a U-shaped function. Data was collected in blocks of 100 trials, consisting of 50 pairs of trials, one from each staircase. The order of the staircases within each pair was varied in accordance with a pseudo-random number generator.

Two subjects were run: one (the author) was an experienced psychophysical observer very much aware of the purpose of the experiment. The other (LL) was a college undergraduate who had had some practice at making psychophysical judgments but was naive with respect to the purpose of the experiment. An attempt was made to run a third subject (an inexperienced college undergraduate), but this subject did not report twinkle at any value of the parameter α between 0 and 1, although he did report seeing two planes of motion for all values. It is possible that he did not completely understand the instructions, or it may reflect large individual differences in sensitivity to noise when masked by coherent motion.

RESULTS

Typical data from a single run of the experiment are shown for two subjects in Fig. 4. In addition to the proportion of "twinkle seen" responses at each value of the parameter α, each graph also shows the number of staircase trials presented, which give a rough indication of the reliability of the proportion data.

The data were analyzed by splitting the resulting data sets in the center portion of the U where twinkle was never seen. (There was always a large dead zone between the two limbs of the U, so there was no problem deciding where to break the data, as there might have been if the bottom of the U had not extended all the way to zero.) Each of the two limbs was then fit with a cumulative Gaussian. Each fit minimized the squared deviations by varying two parameters: the position of the inflection point, and the slope or semi-interquartile distance (SIQD). The fitting procedure was a weighted least squares probit analysis, which is described in detail by Mulligan and MacLeod (1988). This fitting procedure utilized all of the observations (not just the staircase reversals).

Each observer ran four blocks, each of which was fit individually as described above. Figure 5 shows the mean results for the two subjects. The curves were drawn by hand using the mean parameters from the fitting procedure. The fits yielded the horizontal location of the inflection point (which is plotted) and the semi-interquartile distance, which is indicated by the laterally extending bars.

It should be noted that there are two scales given for the x-axis in Fig. 5: the lower scale indicates the variable luminance as a fraction of the (fixed) light and dark luminances in the pattern. The upper scale indicates the absolute luminance (in arbitrary units). Because the luminance of the dark cells was chosen to be exactly one-ninth of that of the brightest cells, the geometric mean of the light and dark values corresponds to one-third of the maximum, which corresponds to a value of 0.25 on the lower scale. It was considered desirable to have a non-zero minimum luminance in order that there would be a well-defined geometric mean, which would be the predicted value of the variable luminance to minimize twinkle in the presence of a logarithmic transformation.

The two small arrows at the bottom of the figure indicate the midpoint of the range spanning the two mean inflection points, i.e. the midpoint of the "pure transparency" zone. These are our estimates of the location of the bottoms of the U-shaped functions. These both lie slightly to the right of 0.5 on the fractional scale (which corresponds to the arithmetic mean of the light and dark luminances). The geometric mean (luminance $\frac{1}{3}$) can be seen to lie at the edge of the noise-free zone.

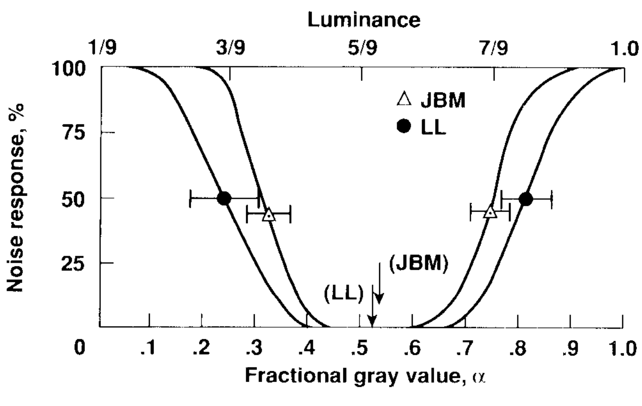

FIGURE 5. Summary data for subjects JBM and LL, similar to Fig. 4. The points plotted represent the mean values of the abscissas of the inflection points of the cumulative Gaussians which were fit to each set of raw data. The lateral "error bars" are not error bars in the usual sense, but show the mean of the semi-interquartile difference of the fit curves. The curves in the figure were drawn by eye to be consistent with the mean parameter values. The small arrows near the center of the figure indicate the mean value of both sets of inflection points (for each subject), providing an estimate of the center of the twinkle-free zone.

DISCUSSION

Unlike the results of Stoner *et al.* (1990), the findings reported in the current study are *not* consistent with smoothest transparency occurring for luminances consistent with actual physical transparency or translucency. Although the range of luminances where no flicker or twinkle is seen is wide, and includes (barely) the case corresponding to multiplicative transparency (shadows) at one extreme, it is clearly more closely centered on the arithmetic mean than on the geometric mean. The range of luminances used by Stoner *et al.* included values corresponding to partial transparency (translucency), which fall between the geometric mean and the lower limit of values in the present experiment. Many of these values produced incoherent component motion in the experiments of Stoner *et al.* but produced significant amounts of visual noise in the present experiments. Stoner *et al.* concluded from their result that the visual system "knows" about transparency, and exploits this knowledge in the interpretation of visual motion. If this were the case, one might expect that this knowledge might also suppress the perception of visual noise when presented with the stimuli from the present experiment which also could have been produced by translucent objects. How can one explain the apparent conflict of the present results with those of Stoner *et al.*? One possibility is that their stimulus (plaids made from thin criss-crossing stripes) contains strong figural cues to the two patterns: the two patterns which are seen as transparent when moved can be easily segmented even in the static case; in fact, a motivated observer can probably cause the *static* pattern to be seen transparently, much in the way that observers can "will" Necker cube reversals. The two components of random-dot stimulus employed in the present experiments, however, cannot be segmented in the absence of motion. This stimulus is more likely, therefore, to tell us something about bottom-up stimulus transformations that occur in the analysis of motion. The results of Stoner *et al.* can be interpreted as powerful evidence of the effects of top-down processing on the resolution of ambiguity in the motion system.

The results of the present study are reminiscent of an earlier result reported by Anstis and Mather (1985). They created an ambiguous apparent motion stimulus by presenting a light bar and a dark bar on a gray background. After a small delay, the positions of the two bars were exchanged. They found that various motion percepts could be evoked depending on the luminance of the gray background. When the background luminance was equal to the luminance of one of the bars, that bar disappeared against the background, and the bar which remained visible was seen jumping back and forth between the two positions. When the background luminance was changed slightly, so that both bars were visible, but one at much greater contrast, the bar with higher contrast remained the one which was seen to move. For a range of intermediate background luminances, however, a neutral percept resulted, in which either no motion was seen, or both bars were seen moving in opposite directions. They called the luminance value in the center of this range the "indifference luminance". Two hypotheses for the value of the indifference luminance were: (1) the indifference luminance would be the arithmetic mean of the light and dark bar luminances, corresponding to *equal luminance changes*; or (2) the indifference luminance would be the geometric mean of the light and dark bar luminances, corresponding to *equal edge contrasts*. They found that the data validated hypothesis 1, i.e. the indifference luminance was the arithmetic mean of the bar luminances, a result similar to that found in the present experiments.

The paradigm introduced by Anstis and Mather has recently been restudied and generalized to color by Shioiri, Cavanagh and Favreau (1989). Unlike Anstis and Mather, Shioiri *et al.* found evidence for a weak compressive nonlinearity, i.e. they measured an indifference luminance slightly less than the arithmetic mean of the test luminances. Their results were still much closer to the arithmetic mean than to the geometric mean however. One possibility is that they were observing effects of the retinal nonlinearity first reported by MacLeod, Williams and Makous (1985), and subsequently studied by Chen, Makous and Williams (1993).

CONCLUSIONS

The results of this study suggest that the appearance of noise or twinkle in addition to motion in complex motion displays is well described by an additive or linear decomposition, such as spatio-temporal Fourier components. Such a decomposition is not what would be required to separate common environmental features like shadows from underlying reflectance patterns. The results suggest a high degree of linearity in the pathway subserving visual motion processing, consistent with that reported by Anstis and Mather (1985).

REFERENCES

Adelson, E. H. (1990). Lightness judgements and perceptual organization. *Investigative Ophthalmology and Visual Science (Suppl.)*, *31*, 265.

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.

Adelson, E. H. & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature (London)*, *300*, 523–525.

Anstis, S. M. & Mather, G. (1985). Effects of luminance and contrast on the direction of ambiguous apparent motion. *Perception*, *14*, 167–179.

Beck, J. (1978). Additive and subtractive color mixture in color transparency. *Perception & Psychophysics*, *23*, 265–267.

Beck, J. (1985). Perception of transparency in man and machine. *Computer Vision, Graphics and Image Processing*, *31*, 127–138.

Beck, J. (1986). The perception of transparency: A reply to Brill. *Computer Vision, Graphics and Image Processing*, *35*, 272–273.

Beck, J. & Ivry, R. (1988). On the role of figural organization in perceptual transparency. *Perception & Psychophysics*, *44*, 585–594.

Beck, J., Prazdny, K. & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, *35*, 407–422.

Brill, M. H. (1984). Physical and informational constraints on the perception of transparency and translucency. *Computer Vision, Graphics and Image Processing*, *28*, 356–362.

Brill, M. H. (1986). Perception of transparency in man and machine: A comment on Beck. *Computer Vision, Graphics and Image Processing*, *35*, 270–271.

Chen, B., Makous, W. & Williams, D. R. (1993). Serial spatial filters in vision. *Vision Research*, *33*, 413–427.

Chen, B. & Makous, W. (1990). Temporal modulation at high spatial frequencies. *Investigative Ophthalmology and Visual Science (Suppl.)*, *31*, 428.

Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (pp. 209–228). Cambridge Mass.: MIT Press.

MacLeod, D. I. A., Williams, D. R. & Makous, W. (1992). A visual nonlinearity fed by single cones. *Vision Research*, *32*, 347–363.

Masin, S. C. (1984). An experimental comparison of three- versus four-surface phenomenal transparency. *Perception & Psychophysics*, *35*, 325–332.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, *13*, 59–66.

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 90–98.

Metelli, F. (1985). Stimulation and perception of transparency. *Psychological Research*, *47*, 185–202.

Metelli, F., DaPos, O. & Cavedon, A. (1985). Balanced and unbalanced, complete and partial transparency. *Perception & Psychophysics*, *38*, 354–366.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1986). The analysis of moving visual patterns. *Experimental Brain Research (Suppl.)*, *11*, 117–151.

Mulligan, J. B. & MacLeod, D. I. A. (1988). Reciprocity between luminance and dot density in the perception of brightness. *Vision Research*, *28*, 503–519.

Shioiri, S., Cavanagh, P. & Favreau, O. E. (1989). Non-linearity in color space measured by apparent motion. *Investigative Ophthalmology and Visual Science (Suppl.)*, *30*, 130.

Stoner, G. R., Albright, T. D. & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature (London)*, *344*, 153–155.

Tudor-Hart, B. (1928). Studies in transparency, form and color. *Psychologische Forschung*, *10*, 255–298.

Watson, A. B. & Ahumada, A. J. Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America, A*, *2*, 322–342.

Acknowledgements—This work was supported by NASA RTOP 505-67. The experiments were first reported at the 1988 annual meeting of the Optical Society of America in Santa Clara, Calif. Thanks to Lee Stone and David Heeger for comments on the manuscript.