Pergamon

0042-6989(95)00195-6

Vision Res., Vol. 36, No. 9, pp. 1291–1310, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

The Interpretation of Visual Motion: Evidence for Surface Segmentation Mechanisms

GENE R. STONER*†, THOMAS D. ALBRIGHT*

Received 16 January 1995; in revised form 26 June 1995

The independent motions of objects in a visual scene are commonly manifest as overlapping retinal motions. A consequence of this overlap is the creation of spurious retinal image features—such as corners and terminated contours—that bear no direct relation to the motions of the objects that give rise to them. To reconstruct object motions, these emergent features must be distinguished from the retinal motions of real object features. This process can be studied using visual stimuli known as plaid patterns, which provide a laboratory archetype for the ubiquitous real-world circumstance of two surfaces with overlapping retinal projections. By adjusting luminance relationships in a plaid pattern it is possible to influence the perceptual interpretation of image features, such that they are seen as either an emergent consequence of occlusion or as real variations in surface reflectance. In the former case, the plaid is most likely to be to perceived as two independently moving surfaces, whereas the latter generally elicits a percept of a single moving surface. This dependence of motion perception on luminance configuration can be viewed as evidence for the involvement of surface segmentation mechanisms, which distinguish between real and emergent image features by promoting a depth-ordered neural representation of surfaces. An alternative interpretation, which does not demand such depth-ordering and feature classification, asserts that the effect of luminance configuration can be accounted for by attendant variations in the distribution of moving Fourier components. To evaluate these two proposed mechanisms, we designed novel plaid stimuli in which surface segmentation cues could be varied independently of changes in the distribution of Fourier components. Perceived motion was found to be highly correlated with the presence of appropriate segmentation cues and uncorrelated with the distribution of Fourier components. These results refute the Fourier components hypothesis, and they support our proposal that surface segmentation plays a critical role in the interpretation of visual motion signals.

Motion Plaid Segmentation Fourier Transparency

INTRODUCTION

Objects lying at different distances from an observer frequently have overlapping projections upon the two-dimensional retinal image. "Image segmentation" refers to the perceptual decomposition of such images into the constituent objects of the visual scene. A fundamental feature of image segmentation is the existence of a multi-valued representation of one or more scene attributes (e.g. color, depth, motion) associated with a single spatial location in the retinal image. Absence of segmentation, on the other hand, implies that there are only single-valued representations for each image location.

Two-dimensional plaid patterns (De Valois *et al.*, 1979; Adelson & Movshon, 1982) are constructed by superimposing two one-dimensional gratings (Fig. 1). Beginning with the work of Wallach (1935), this class of stimuli has played a prominent role in illuminating the mechanisms underlying segmentation in the domain of motion processing. "Coherent motion" is the term that has been used to describe the condition in which the two gratings that compose a drifting plaid pattern are seen to move as a single surface. This percept can be interpreted as evidence for a single-valued or non-segmented neuronal representation. Conversely, "non-coherent motion" describes the condition in which the component gratings are perceived to move independently; a state that implies an independent representation of each grating's motion. Neurophysiological data obtained using plaid patterns as visual stimuli also support the existence of both segmented and non-segmented neuronal representations (Movshon *et al.*, 1985; Rodman & Albright, 1989; Stoner & Albright, 1992a).

One of the most potent and well-documented segmentation cues in static images is luminance configuration (LC) (e.g. Metelli, 1974; Beck *et al.*, 1984). As shown in Fig. 2, LCs consistent with the superimposition of

*Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, U.S.A.

†To whom all correspondence should be addressed at: The Salk Institute, P.O. Box 85800, San Diego, CA 92138, U.S.A. [*Fax* +1 619 552 8285].

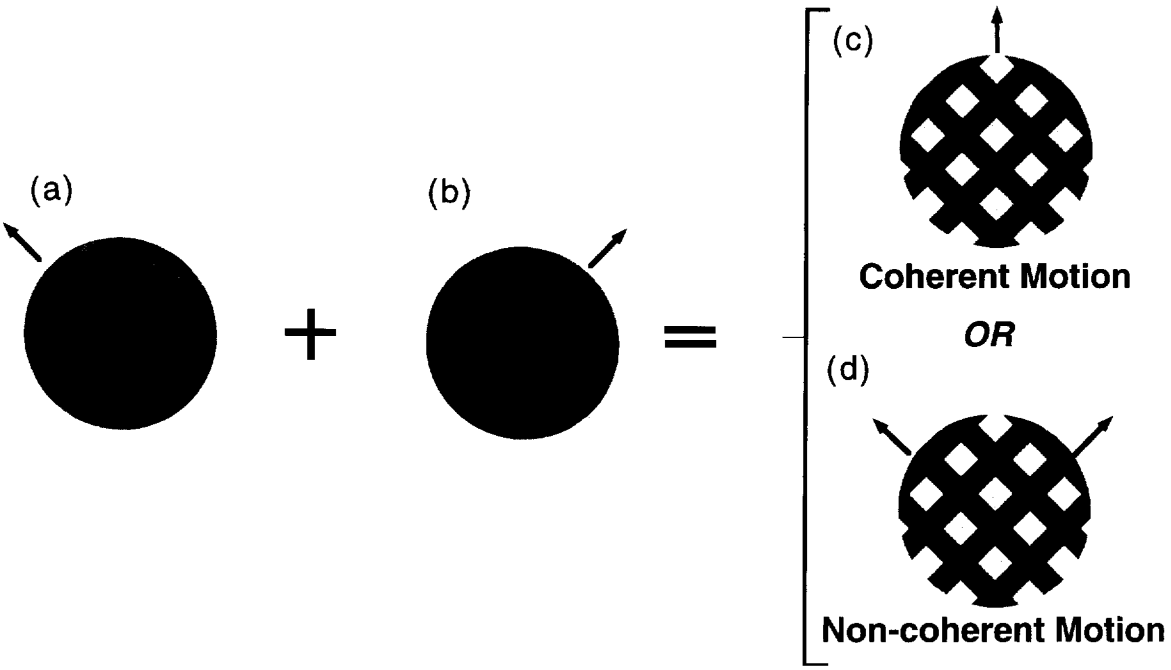

FIGURE 1. Conventional moving plaid patterns are produced by additive superimposition of two drifting periodic gratings (a, b). The resultant percept is either that of a coherently moving two-dimensional pattern (c) or two one-dimensional gratings sliding past one another (d), depending on a variety of stimulus parameters.

transparent or opaque surfaces*—in accordance with the projective addition of the light reflected from such surfaces—are generally sufficient to elicit the corresponding percept. To directly investigate the role of surface segmentation in motion coherency, Stoner *et al.* (1990) exploited LC as a means to manipulate perceptual transparency in moving plaid patterns (Fig. 3). Human observers were found most likely to report motion non-coherency when LC was adjusted to elicit a percept of overlapping surfaces. In concert with related studies (e.g. Shimojo *et al.*, 1989; Trueswell & Hayhoe, 1993; Vallortigara & Bressan, 1991; Kersten *et al.*, 1992; Kooi *et al.*, 1992; Adelson & Movshon, 1982), these data suggest that information concerning the depth-ordering of superimposed image features is used to "gate" the integration of visual motion signals (Albright & Stoner, 1995).

Challenging the role of surface segmentation: The Fourier components hypothesis

An alternative explanation for the results of Stoner *et al.* (1990) follows from the fact that LC adjustments necessarily alter the spectrum of Fourier components associated with the moving plaid (e.g. Stoner & Albright, 1992b; Trueswell & Hayhoe, 1993; Mulligan, 1993; Noest & van den Berg, 1993; Plummer & Ramachandran, 1993). Plaids that mimic transparent or opaque occlusion possess Fourier components that move in the coherent pattern direction (see Appendix and Fig. 4). This observation leads to the prediction, contrary to the findings of Stoner *et al.* (1990), that coherent motion should be *more* likely for transparently configured plaids than for many of their non-transparent counterparts. If, however, one embraces the commonly accepted idea that image intensity is (approximately) logarithmically compressed by an early stage of neuronal processing (see MacLeod, 1978; MacLeod *et al.*, 1992), it can be shown that Fourier components moving in the pattern direction are minimized when plaid stimuli are constructed as multiplicatively attenuating transparent surfaces. With this qualification, the predictions of a mechanism based upon spatio-temporal Fourier energy are roughly consistent with the psychophysical results of Stoner *et al.* (1990). This spatio-temporal Fourier energy explanation (henceforth simply termed the "Fourier components hypothesis") thus reclaims the effect of LC on motion coherency from the realm of surface segmentation processes and reframes it in more conventional terms. As we will show, however, this explanation suffers from the lack of generality it affords.

Dependence of perceptual transparency upon foreground/background assignment: Insufficiency of the Fourier components hypothesis

The Fourier components hypothesis predicts that non-coherency should occur for a fixed LC regardless of

*The terms "transparency", "opacity", and "occlusion" have been used in a variety of contexts (including colloquial) and their precise meanings are not always clear. Here we use occlusion to refer to the condition in which a proximal (foreground) surface overlaps a distal (background) surface in the formation of the retinal image. Transparent and opaque are terms we use to refer to the physical properties of occlusive surfaces. Specifically, transparent surfaces partially attenuate light reflected off of the surfaces they occlude; opaque surfaces provide complete attenuation.

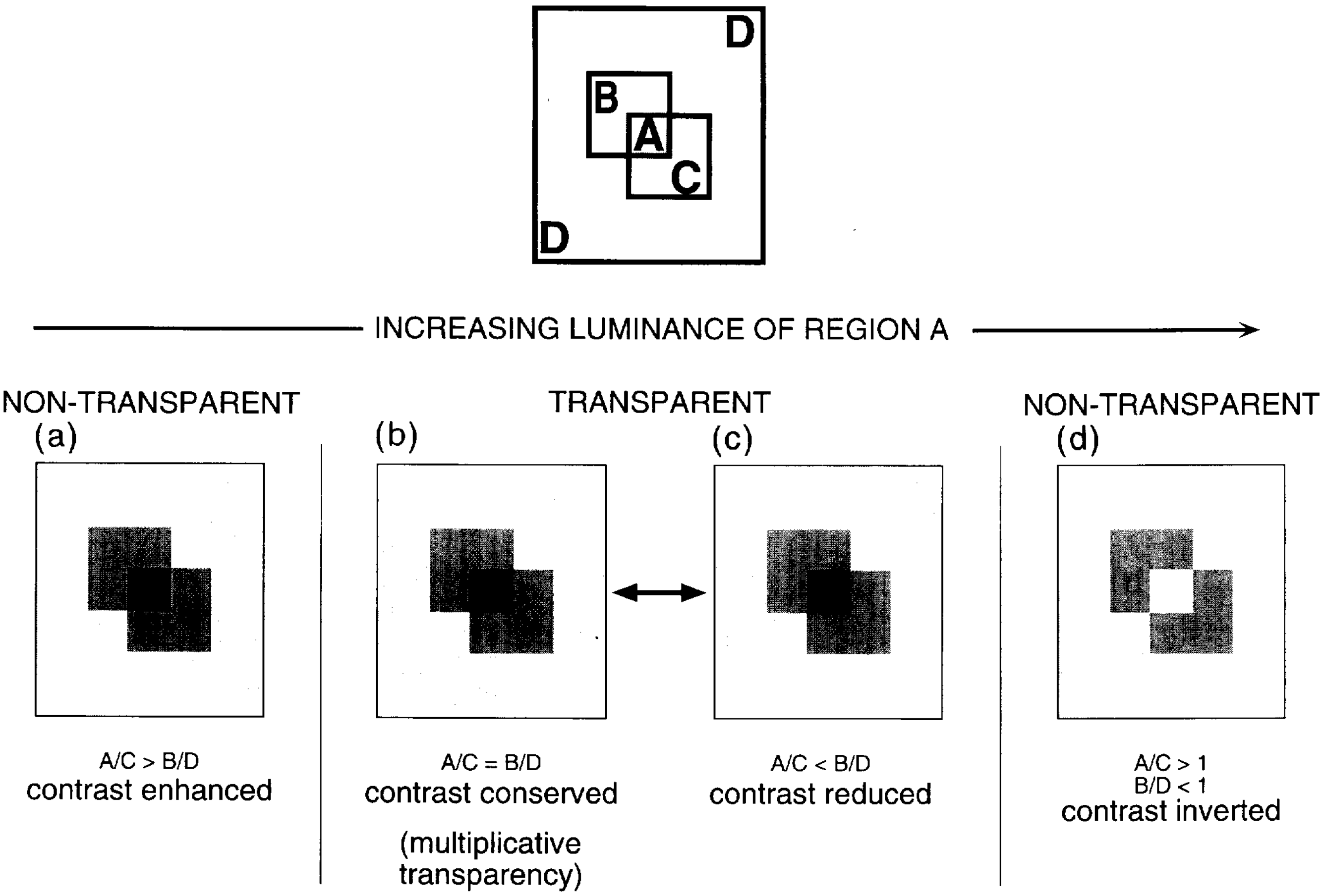

FIGURE 2. The luminance rules governing perceptual transparency are derived from the physics of transparency (Metelli, 1974; Beck *et al.*, 1984). Simply put, luminance ratios within the pattern must be physically consistent with the transmittance of light from a far surface *through* a near surface. Appropriate luminance ratios convey a sense of depth ordering (and hence image segmentation) in a pattern devoid of other depth cues. In this "cascading rectangles" example, the rectangle designated C (or B) can be interpreted as a transparent or opaque occluder partially overlying rectangle B (or C). The rectangle designated A indicates the region of overlap, and the larger rectangle D is background. The luminance relationships amongst regions A, B, C, and D determine perceptual transparency, in accordance with the physical constraints of surface transmittance and reflectance. In practice, these relationships can be critically altered by manipulating the brightness of a single subregion, such as A (the region of overlap). *Perceptual transparency* is most likely if the luminance contrast of the background viewed through the apparent foreground surface (i.e., the contrast ratio A / C) is conserved (b) or reduced (c) relative to the contrast of the unobscured background surfaces (B / D). By contrast, neither transparent nor opaque occlusion is likely if the luminance contrast viewed through the putative foreground region is either enhanced (a) or inverted (b). See text for details.

figural interpretation (Noest & van den Berg, 1993; see also Appendix). This prediction is directly at odds with our proposal that surface segmentation processes involved in perceptual transparency are involved in the segmentation of dynamic surfaces. The reason for this assertion follows from a corollary of the rules governing perceptual transparency: luminance contrast within a background surface is normally diluted when viewed through a transparent (or opaque, in the extreme) foreground surface. This corollary is a simple physical consequence of light mixture, and it accounts for the fact that perceptual transparency is most likely when the contrast of the background viewed through the transparent foreground surface is less than or equal to the contrast of the background viewed directly (Fig. 2) (Metelli, 1974; Beck *et al.*, 1984). It follows that a given LC should elicit perceptual transparency only if particular sub-regions of the pattern are interpreted by the viewer as foreground and other regions as background. Furthermore, if motion integration is informed by the same segmentation mechanisms, motion coherency should also be dependent upon foreground/background (F/B) interpretation.

There are a variety of image cues that influence F/B interpretation. An otherwise appropriate LC will generally fail to elicit a percept of transparency if the F/B interpretation it supports is inconsistent with that promoted by another cue. We can exploit this fact to force a disassociation between the predictions of the image segmentation and Fourier components hypotheses. If the effects of LC on motion coherency truly reflect the involvement of surface segmentation mechanisms, the LCs that generate peak non-coherency should depend upon cues for F/B interpretation that are not linked to critical Fourier components. Conversely, if the Fourier components hypothesis has general validity, peak non-coherency should be associated with a fixed LC (yielding

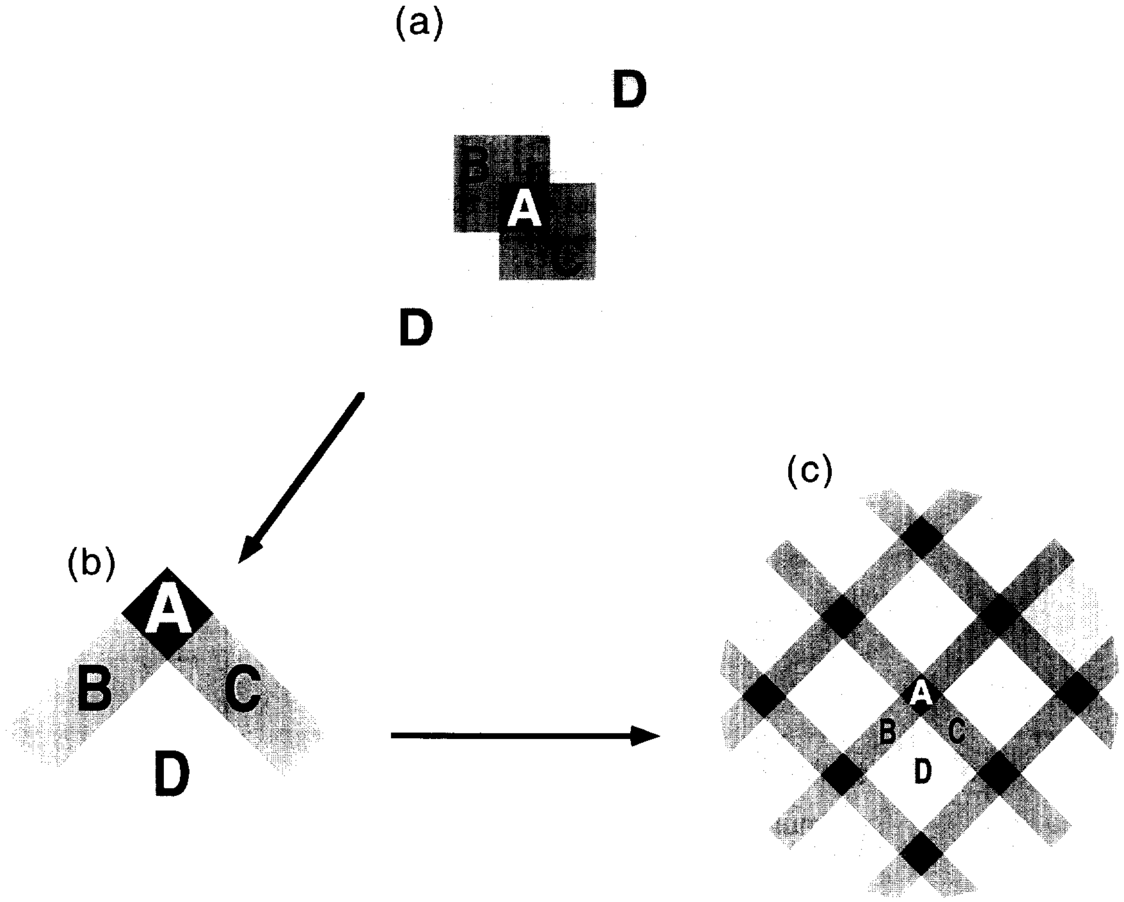

FIGURE 3. The procedure for manipulating transparency in plaid patterns. Each plaid (c) can be viewed as a tesselated image composed of four distinct repeating subregions (b), identified as A, B, C, and D. These plaid subregions bear precisely the same relationship to one another as in the cascading rectangle configuration (a): region D is normally seen as background (owing to its relatively larger size). Regions B and C are seen as narrow overlapping surfaces and the remaining region A as their intersection. Perceptual transparency is manipulated by varying the luminance of region A, while the luminances of regions B, C, and D are held constant.

appropriate components of Fourier energy), *regardless of figural interpretation* (Noest & van den Berg, 1993).

To evaluate these two hypotheses, we devised three experiments. The goal of Expt I was to systematically establish the pairings of F/B assignment and LC that generate perceptual transparency in *static* stimuli. This knowledge permitted us to look, in Expt II, for evidence of surface segmentation's influence in the processing of dynamic images. In Expt III we utilized moving stimuli that lacked any consistent cue for F/B interpretation. These stimuli exhibited a Necker cube-like perceptual metastability, such that F/B interpretation varied on a trial-by-trial basis. This F/B metastability, in the presence of an invariant retinal image (and, hence, in the absence of any change in spatio-temporal Fourier components), provided the most definitive means to evaluate our two competing hypotheses. In addition to these experimental procedures, we subjected all of our plaid stimuli to Fourier decomposition, in order to determine the strength of Fourier components moving in pattern and component directions. This analysis indicated those stimuli that, according to the Fourier components hypothesis, should be most likely to yield maximal non-coherency.

Experiments I and II confirmed that perceptual transparency (for static patterns) and motion coherence (for moving plaids) exhibit a parallel and dramatic dependence upon agreement between cues for LC and F/B interpretation. Experiment III revealed, furthermore, that physically identical stimuli could produce either coherency or non-coherency, in a manner that was highly correlated with F/B interpretation. This critical role of F/B assignment refutes the Fourier components hypothesis and it supports our assertion that the motion system has access to image segmentation processes that incorporate tacit knowledge of the rules governing retinal image formation from natural scenes.

GENERAL METHOD

Subjects

Five human subjects participated in these experiments. Three (GN, RC, and CM) were completely naive with regard to the goals of the experiments, whereas one subject (GB) was knowledgeable of basic issues, but somewhat skeptical of the guiding hypotheses. The fifth

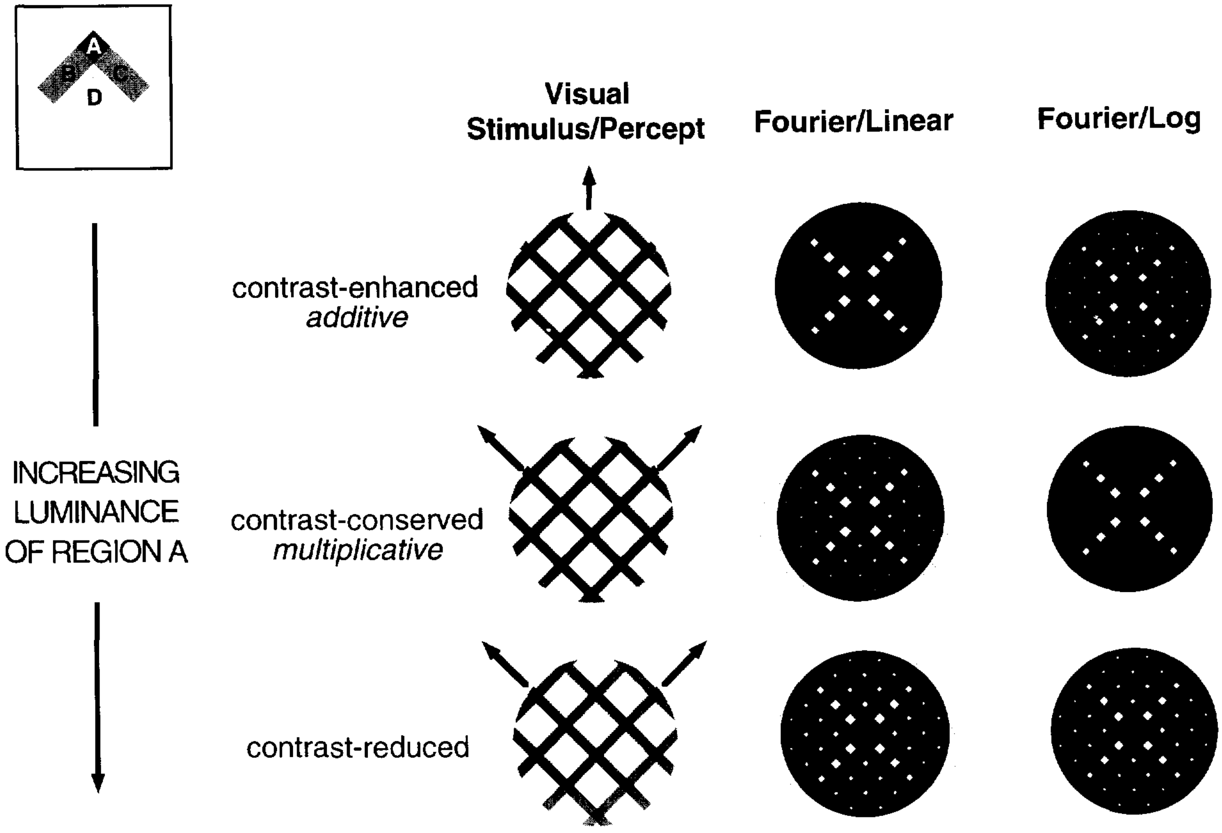

FIGURE 4. Fourier components explanation for the effects of luminance cues for transparency on perceptual motion coherency observed by Stoner *et al.* (1990). The inset box at the upper left indicates label designations for plaid subregions. Left column: luminance configuration for three plaid patterns used in the experiments of Stoner *et al.* These plaids differed only in the luminance of the small rectangular regions designated A, which were typically perceived as foreground under these conditions. The "contrast-enhanced plaid" at the top was formed such that the luminance of region A was determined by simple linear addition of the two components (regions B and C). This configuration is physically *in*compatible with transparency and subjects typically perceived coherent pattern motion (upward arrow). The "contrast-conserved plaid" in the center was formed such that the luminance of region A was determined by multiplicative combination of the two components. This configuration is compatible with transparency and subjects typically perceived the independent motions of the two component gratings (diagonal arrows). The "contrast-reduced plaid" at the bottom was formed such that the luminance of region A was equal to another non-linear combination of component intensities (A ~ = B). This configuration is also consistent with transparency (see Fig. 2) and subjects usually report non-coherent pattern motion (diagonal arrows) for it as well. Center column: Fourier amplitude spectra associated with the three plaid patterns shown on the left. Fourier spectra were determined using the *Mathematica* FFT utility on a Macintosh IIcx. The zero-frequency point lies in the center of each plot. The location of each small rectangle in the spectral plots is indicative of frequency (radius) and orientation (polar angle); the size indicates the amplitude of the associated frequency. Scaling (x-frequency, y-frequency, and amplitude) has been arbitrarily chosen and is the same for all plots in this and other figures; only relative amplitudes of Fourier components associated with grating and pattern directions are of interest for this analysis. For convenience of exposition and interpretation, Fourier components associated with *grating* motions are evidenced by diagonal columns of dots that pass through the zero point. Fourier components associated with the two-dimensional *pattern* motion are evidenced by a vertical (or horizontal) column of dots passing through the zero point. The additive plaid (top row) is identified with the Fourier "null-point", as it contains only those components associated with the independent grating motions. To conform with the luminance configuration rules for perceptual transparency (outlined in Fig. 2), the luminance of the region of foreground overlap (region A) must be increased beyond the value associated with additive superimposition. As a consequence, the contrast conserving (center row) and contrast reducing (bottom row) plaids both contain Fourier components that move in the pattern direction. The "Fourier components hypothesis" asserts that low-level motion detectors are sensitive to these emergent pattern components, and that perceptual motion coherency is simply determined from the strength and directional distribution of Fourier components in the plaid (see text and Appendix for details). The Fourier components hypothesis thus predicts that maximal non-coherency should be elicited by the additive plaid. This is contrary to the results of Stoner *et al.* (1990). Right column: Fourier amplitude spectra associated with the three plaid patterns shown on the left, following point-wise logarithmic compression of image intensities. This non-linear transformation introduces pattern Fourier components to the additive case (top row), and the Fourier null-point is now associated with the multiplicative plaid (center row). The distribution of Fourier components following log-compression predicts maximal non-coherency for this multiplicative condition, which is roughly consistent with the results of Stoner *et al.* (1990).

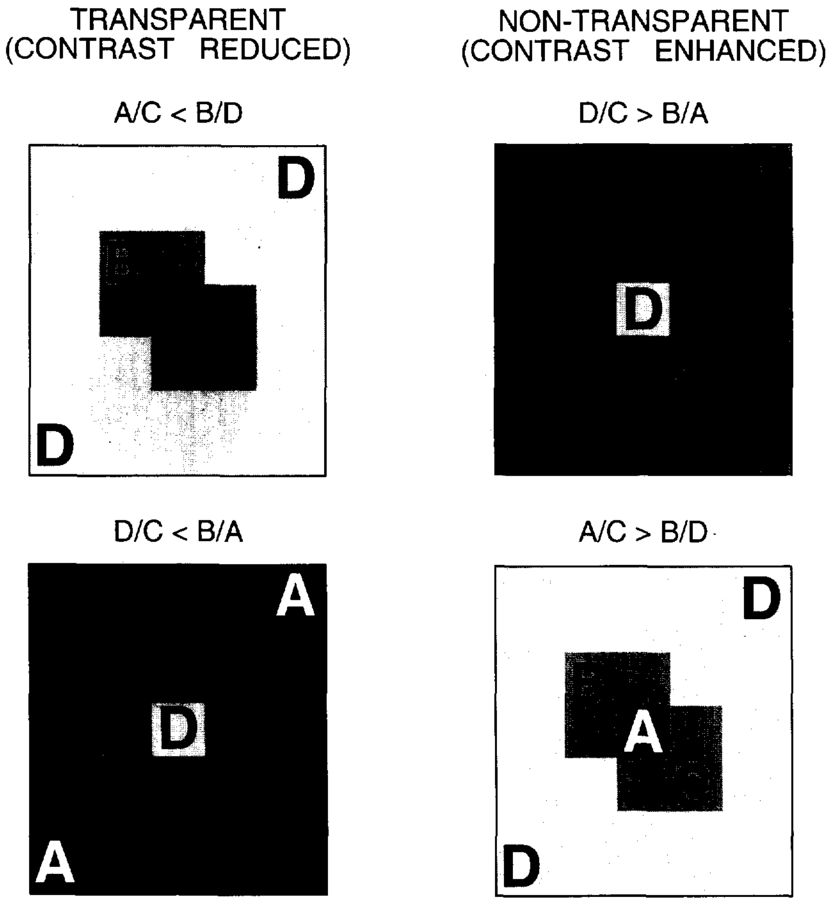

FIGURE 5. Schematic illustration of means used in Expt I to manipulate perceptual transparency for *static* cascading rectangle patterns. Perceptual transparency is known to be influenced by the relative luminances of the four sub-regions of the pattern (i.e. the "luminance configuration") (e.g. Metelli, 1974; Beck *et al.*, 1984). Luminance configuration does not, however, uniquely determine transparency: we predicted that transparency should also depend upon F/B assignment of the subregions in the pattern. Luminance configuration was manipulated by altering the luminance of region A, while keeping the luminances of regions B, C, and D constant. Reversal of F/B assignment was accomplished by simply switching the luminance assignments of the small and large regions (regions A and D). The effect of this F/B switch on perceptual transparency is shown for two different luminance configurations (top and bottom rows—corresponding to the contrast enhanced and contrast reduced LCs of Fig. 4). The predicted perceptual effects of these manipulations can be interpreted by evaluating their consistency with the following hypothesis: the large framing rectangle and one of the smaller rectangles are background surfaces; the other smaller rectangle is a foreground surface, and the small central square is the region of overlap between foreground and background rectangles. In order for this hypothesis to be true, the contrast between the unoccluded background surfaces must always be greater than or equal to that between the same surfaces as *viewed through* the foreground surface. Top-left panel: the luminance configuration is consistent with background surfaces B and D being viewed through foreground surface C. *Prediction*: transparent. Top-right panel: the relative intensities of the four subregions are identical to those in the top-left, but the assignments of regions A and D have been reversed. Doing so causes an enhancement of background contrast when viewed through the putative foreground region—a condition that is not likely to be realized under natural conditions. *Prediction*: non-transparent. Bottom-left panel: the luminance configuration has been altered relative to the top row: region A is less bright. Although the luminance and spatial relationships amongst regions B, C, and D are identical to those in the top-right, this arrangement is consistent with background surfaces B and D being viewed through foreground surface C. *Prediction*: transparent. Bottom-right panel: the relative intensities of the four subregions are identical to those in the bottom-left, but the assignments of regions A and D have been reversed. Doing so causes an enhancement of background contrast when viewed through the putative foreground region—a condition that is physically improbable. *Prediction*: Non-transparent.

subject (GS) is one of the authors. All subjects had normal or corrected-to-normal vision.

Visual stimuli

All stimuli were generated using a high-resolution graphics display controller (Pepper SGT, Number Nine Computer Corporation: 640 × 480 pixels, 8 bits/pixel, 60 Hz, non-interlaced) operating in a microcomputer. Stimuli were displayed on a 14 in. analog RGB video monitor (Zenith ZCM-1490, flat technology CRT). Photometric linearization tables were computed and used to reform the non-linear voltage–luminance relationship. The stimuli used in Expts I, II, and III differed only in terms of geometry (i.e. the size, position, and luminance of the different sub-regions) and whether they moved (for coherency judgments, in Expts II and III) or were static (for transparency judgments, in Expt I).

The geometry of all of our stimuli can be described by reference to four "luminance regions", designated A, B, C, and D (Fig. 3). Each stimulus is uniquely defined by a

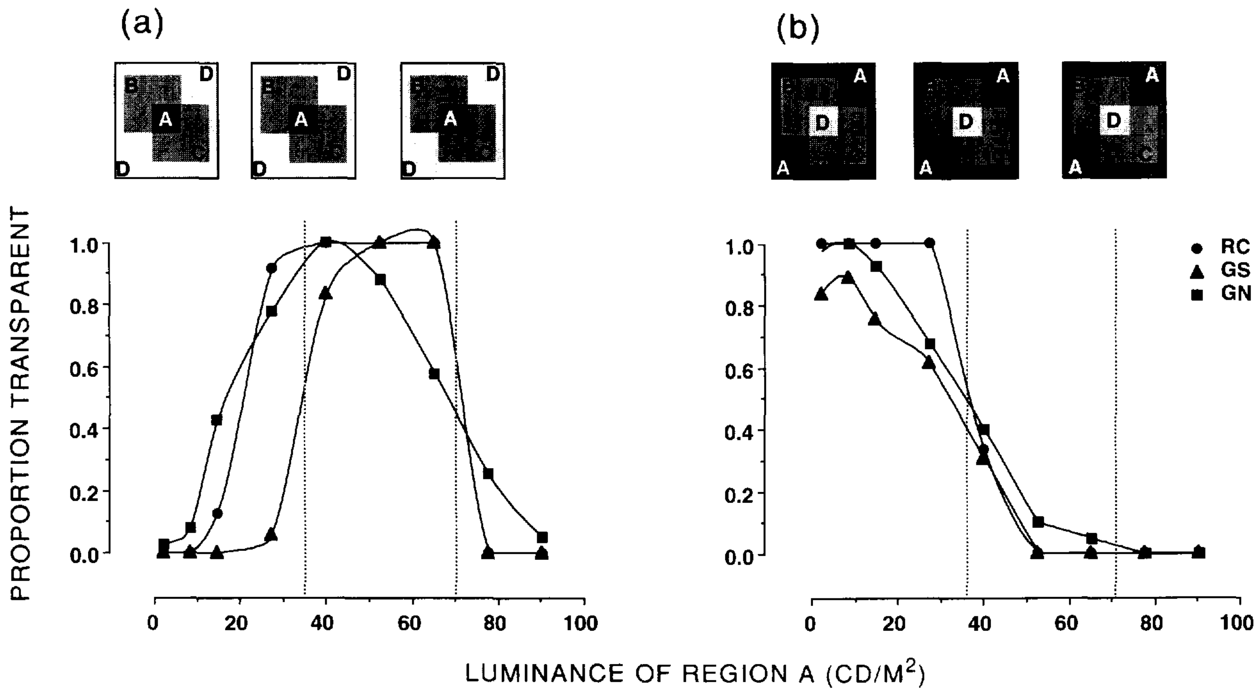

FIGURE 6. The conjoint effects of F/B interpretation and luminance configuration on transparency judgments for the static cascading rectangle patterns used in Expt I (see Fig. 5 for stimulus description). Two independent variables were manipulated: (i) the luminance of the region designated A, which is represented along the abscissa for each graph; (ii) the positional assignment of luminance-varying region A. For the graph in (a), region A is foreground (more precisely, A corresponds to the intersection of two surfaces, B and C, one of which is foreground); for the graph in (b), region A is background. All other stimulus properties, including the luminances of regions B, C, and D, were constant. The approximate appearance of a subset of the different stimulus conditions is illustrated by the icons above each graph. Following each brief (1.33 sec) presentation, subjects were required to report whether the stimulus appeared as two overlapping rectangles or as a single surface possessing variations in surface reflectance. Data are shown for three subjects. The left-most vertical dotted line on each graph indicates the condition in which the luminance of region $A = B * (C / D)$, i.e. the multiplicative transparency condition. The right-most vertical line indicates the condition in which the luminance of region $A = B = C$, i.e. the opaque occlusion condition. When region A is foreground (a), these vertical lines bound the range of region A luminances that are physically compatible with transparency. Within this range, subjects are far more likely to report a percept of transparency. When region A is background (b), the range in which luminances are physically compatible with transparency is lower-bounded by $A = 0$ and upper-bounded by $A = B * (C / D)$. Within this range, subjects are far more likely to report a percept of transparency. Thus, although the range of $A : B : C : D$ luminance relationships is identical for the *A-as-foreground* and *A-as-background* conditions, there is a pronounced shift in the luminance conditions that elicit a percept of transparency. This shift parallels physical compatibility with transparency. Each data point is based on 40 trials. See text for details.

combination of two luminance region properties: (i) the absolute luminance assignment for each region (the "luminance configuration"), and (ii) the position/size assignment for each region. Within each experiment, stimuli were made to differ along these two dimensions by the following means. (i) The LC was altered by changing the absolute luminance of region A; the luminances of regions B, C, and D were constant. (ii) The position/size assignment was altered for regions A and D.

For both static and moving stimuli (Figs 2 and 3, respectively), the luminances of regions B, C, and D were held constant at values of 73, 73, and 143 cd/m^2, respectively. Region A varied from 2 to 90 cd/m^2 in approximately equal increments of 9 cd/m^2, rendering a total of nine stimulus conditions. Stimuli were viewed from a distance of 57.3 cm. The ambient light level in the experimental room was approx. 2 cd/m^2 and the mean luminance of the screen during the inter-trial interval was 57 cd/m^2. A chin rest was used for head stabilization.

Psychophysical procedure

Excepting the type of judgment made (transparent vs non-transparent or coherent vs non-coherent), the procedures for Expts I–III were identical, as follows. Data were collected using a two-alternative, forced-choice procedure. Each subject was instructed to fixate a small spot at the center of the display for the duration of each trial. Trials were initiated by a key-press once the subject attained fixation. Stimulus duration was 1.33 sec. Subjects were required to indicate their dominant percept (transparent vs non-transparent or coherent vs non-coherent motion) with an appropriate key-press at the end of each trial. Each subject was initially presented with a series of practice trials, which was continued until performance become stable and the subject expressed confidence and a clear understanding of the requirements of the task. Trials were presented in blocks (36 or 72 trials/block) and the position/size assignment of each stimulus region remained constant within each block. The luminance of region A was varied from trial-to-trial within each block on a pseudo-random schedule.

Regardless of stimulus geometry, each block was composed of an equal number of trials for each LC. With the exception of subjects GB and CM, who completed 42 trials per condition for Expt III, final data analysis was based upon 40 trials per stimulus condition.

EXPERIMENT I: THE EFFECTS OF FOREGROUND/BACKGROUND INTERPRETATION ON PERCEPTUAL TRANSPARENCY IN STATIC DISPLAYS

Although the luminance relationships for transparency established by previous studies (Metelli, 1974; Beck *et al.*, 1984) imply that the probability of perceiving transparency depends upon both the local LC and other cues for F/B interpretation, we know of no explicit investigation of this interdependency. In Expt I, we sought to verify these implications, as they provide a logical foundation for the experiments that follow.

Method

Visual stimuli and psychophysical procedure. Stimuli similar to those depicted in Fig. 2 were presented and subjects were instructed to report whether or not the two rectangular regions labeled B and C appeared as overlapping transparent surfaces. The spatial extent of the large rectangular background region (labeled D in Fig. 2) was 13 deg^2 and each of the smaller rectangles (labeled B and C, and *including* the region of overlap labeled A) was 3 deg^2. The region of overlap itself (A) was 1.5 deg^2.

Two stimulus properties were manipulated as independent variables: the luminance of region A and the position/size assignments of regions A and D. The luminance of region A took those values indicated in the General Method and was varied from trial to trial on a random schedule. The luminances of all other subregions of the pattern were held constant at the indicated values. As illustrated in Fig. 5, manipulation of the position/size assignment of regions A and D was effected by a simple exchange. The purpose of this manipulation was to alter the F/B interpretation of luminance regions A and D. Specifically, because the large enclosive rectangle and the small central rectangle are normally interpreted as background and foreground respectively, the luminance assignments to these regions defines the F/B interpretation of A and D. Other stimulus and procedural details are provided above under General Method.

Results

Informal questioning of subjects confirmed that the large enclosive region was seen as a background surface upon which lay a foreground figure(s). To be consistent with the established rules of perceptual transparency, decomposition of the two overlapping squares into distinct surfaces should be most likely when the luminance contrast of the background, viewed through the putative foreground surface, is attenuated relative to that of the unobscured background. Congruent with these expectations, all subjects reported transparency over a limited range of luminance values. These values were similar across subjects and changed dramatically as a function of the F/B manipulation.

Results obtained from trials in which region A was smaller than region D (i.e. the luminance of the *smaller* region was varied) are shown in Fig. 6(a). Under this condition, the pattern was only physically consistent with two overlapping squares of uniform reflectance if the luminance contrast between A and C was less than or equal to that of the background contrast (expressed as B/D). Consistent with these physical constraints, the peaks of the transparency judgment curves were centered within the zone bounded by opaque and multiplicative transparency [indicated, respectively, by the right-most and left-most vertical dotted lines in Fig. 6(a)]. The peaks of these curves therefore corresponded to conditions in which the luminance of region A was *brighter* than that associated with the multiplicative transparency condition (the condition in which luminance contrast of the background is conserved).

Results obtained from trials in which region A was larger than region D (i.e. the luminance of the *larger* region was varied) are shown in Fig. 6(b). This reversal of the F/B assignments of A and D yielded quite different impressions of transparency. Under these conditions, the peaks in the transparency judgment curves occurred for those conditions in which the luminance of region A was *dimmer* than associated with the multiplicative transparency condition. Thus, as suggested by the work of Metelli, Beck, and others, reports of perceptual transparency were most likely when stimuli satisfied the constraint that the luminance contrast viewed through the foreground transparent surface was less than that of the unobscured background.

By establishing precisely the combination of luminance values and F/B assignment that generate transparency, Expt I constituted a crucial step towards determining whether motion non-coherency is governed by the same stimulus information and segmentation mechanisms underlying the perception of transparency in static displays. The latter was the goal of Expt II.

EXPERIMENT II: FOREGROUND/BACKGROUND INTERPRETATION AND MOTION COHERENCY: THE ROLE OF DUTY CYCLE

The "enclosure" cue for F/B assignment utilized in Expt I (Fig. 5) cannot be applied to moving plaid patterns: the boundaries of the four plaid sub-regions must be shared (see Fig. 3) to prevent the introduction of image features (e.g. corners) that provide unambiguous two-dimensional motion information. Relative surface area is an excellent alternative F/B cue; larger regions are generally seen as background (Wallach, 1935; Koffka, 1935; Petter, 1956). This perceptual effect may reflect a principle of parsimony in the assignment of foreground to regions having common boundaries. In any event, it is a simple matter to exploit size cues in the construction of plaid patterns by adjusting the duty cycle of the component gratings. This was precisely the means by which we ensured a stable F/B interpretation in our

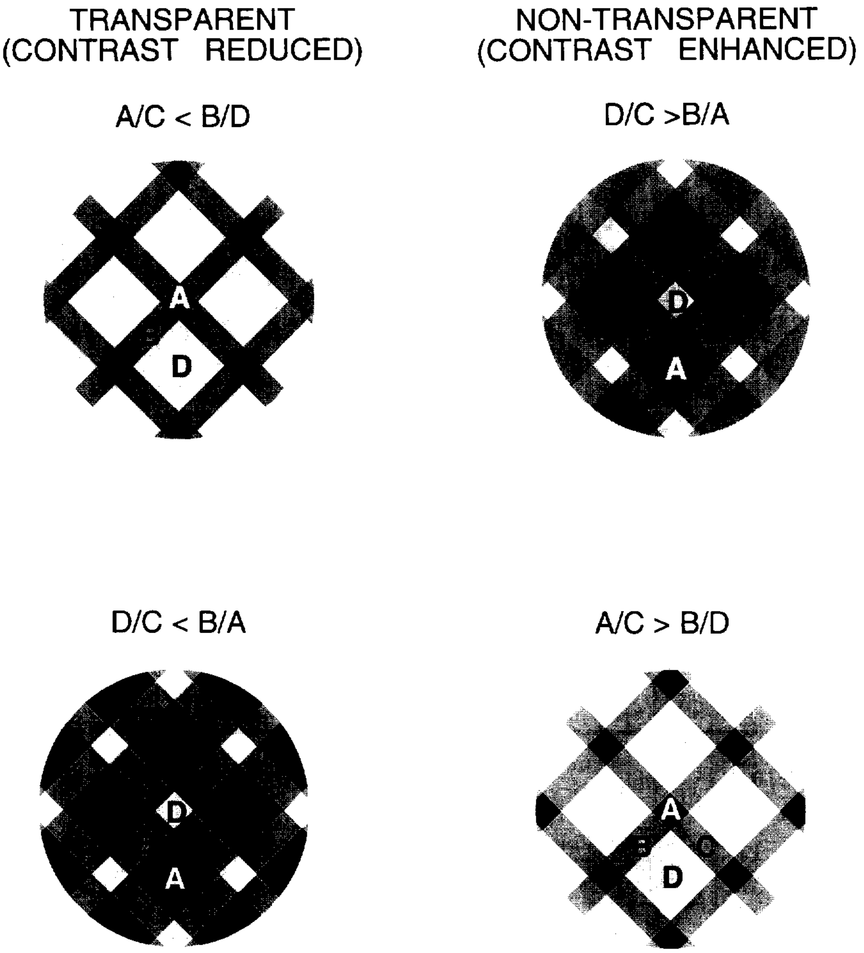

FIGURE 7. Schematic depiction of stimulus configurations used in Expt II to manipulate perceptual transparency and motion coherency for *moving* plaid patterns. As for the static case (Fig. 5), transparency was predicted to be conjointly influenced by luminance configuration and F/B assignment of the subregions of the pattern. Luminance configuration was manipulated by altering the luminance of region A, while keeping the luminances of regions B, C, and D constant. Reversal of F/B assignment was accomplished by simply switching the luminance assignments of the small and large regions (regions A and D). The larger region is typically seen as part of the background (Koffka, 1935; Petter, 1956), and the smaller region as part of the foreground. The figural relationships between the pattern subregions are precisely analogous to the static case, and the same predictive rules should apply. Specifically, the predicted perceptual effects can be interpreted by evaluating their consistency with the following hypothesis: the large rectangle and one set of oblique stripes are background surfaces; the other set of oblique stripes is a foreground surface and the small rectangle is the region of overlap between foreground and background stripes. In order for this hypothesis to be true, the contrast between the unoccluded background surfaces must always be greater than or equal to that between the same surfaces as *viewed through* the foreground surface. Top-left panel: the luminance configuration is consistent with background surfaces B and D being viewed through foreground surface C. *Prediction of surface segmentation hypothesis*: non-coherent motion. *Prediction of Fourier components hypothesis*: coherent motion. Top-right panel: the relative intensities of the four subregions are identical to those in the top-left, but the assignments of regions A and D have been reversed. Doing so causes an enhancement of background contrast when viewed through the putative foreground region—a condition that is physically improbable under natural conditions. *Prediction of surface segmentation hypothesis*: coherent motion. *Prediction of Fourier components hypothesis*: coherent motion. Bottom-left panel: the luminance configuration has been altered relative to the top row: region A is less bright. Although the luminance and spatial relationships amongst regions B, C, and D are identical to those in the top-right, this arrangement is consistent with background surfaces B and D being viewed through foreground surface C. *Prediction of surface segmentation hypothesis*: non-coherent motion. *Prediction of Fourier components hypothesis*: coherent motion. Bottom-right panel: the relative intensities of the four subregions are identical to those in the bottom-left, but the assignments of regions A and D have been reversed. Doing so causes an enhancement of background contrast when viewed through the putative foreground region—a condition that is physically improbable. *Prediction of surface segmentation hypothesis*: coherent motion. *Prediction of Fourier components hypothesis*: coherent motion.

previous study of transparency and motion coherency (Stoner *et al.*, 1990). In those experiments, the asymmetric duty cycle biased F/B assignment such that the plaid sub-region of variable luminance (region A, in Fig. 3) was usually perceived as foreground. The present experiment relied upon manipulations of duty cycle to systematically alter F/B assignment (Fig. 7).

To fulfil our objectives, it is crucial that duty cycle manipulations lead to different predictions from the image segmentation and Fourier components hypotheses. Verifying this entails analysis of the Fourier spectral content of the stimuli. Accordingly, we applied two-dimensional Fourier analysis to static versions of the plaid stimuli used in Expt II (Fig. 8). This analysis, in

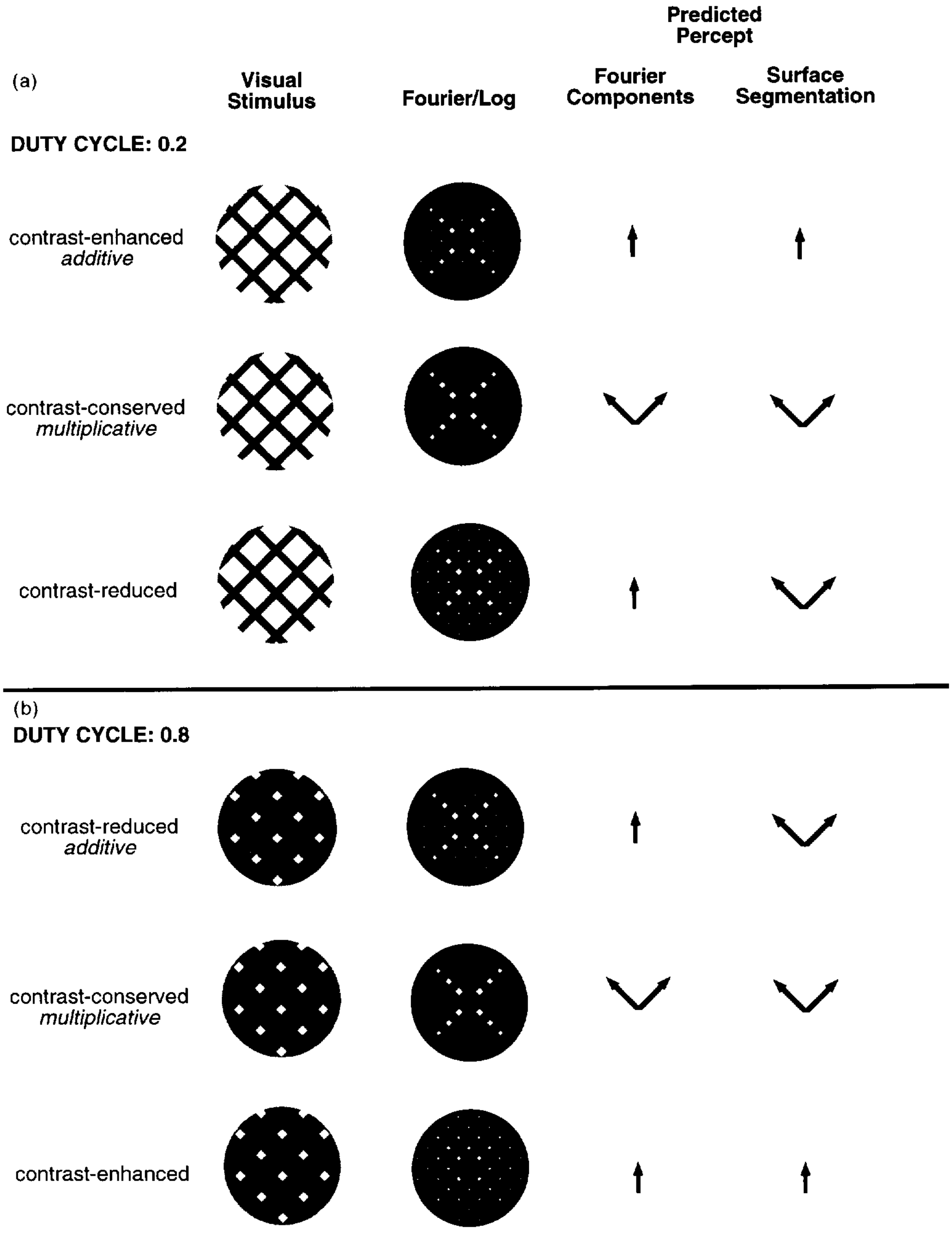

FIGURE 8. Effects of duty cycle manipulations on the Fourier spectra of plaids used in Expt II, illustrating the differential predictions of the "Fourier components" and "surface segmentation" hypotheses. Visual stimuli and accompanying Fourier spectra are shown for three different luminance configurations (rows) and two different duty cycles (a, b). Fourier spectra are illustrated only for *log-compressed* image intensities, for reasons identified in text and Fig. 4. See legend to Fig. 4 regarding plotting conventions for Fourier spectra. (a) Plaids, Fourier spectra, and predictions of the two hypotheses for 0.2 duty cycle condition. Three luminance configurations are shown—contrast-enhanced (additive) (top), contrast-conserved (multiplicative) (center) and contrast-reduced (bottom). The Fourier "null-point" for pattern motion is associated with plaids constructed by multiplicative superimposition. The Fourier components hypothesis therefore predicts maximal motion non-coherence for this condition (indicated by the oblique arrows) with relatively low levels of non-coherence for the other two conditions (indicated by the upward pointing arrows). Both the contrast-conserved and contrast-reduced configurations meet the physical requirements for transparency. Accordingly, the surface segmentation hypothesis predicts that these two configurations will exhibit more non-coherency than the contrast-enhanced additive case. The results of Stoner *et al.* (1990) do not unequivocally distinguish between these two predictions. (b) Plaids, Fourier spectra, and predictions of the two hypotheses for the 0.8 duty cycle condition. While there are small spectral variations that accompany the duty cycle reversal, the Fourier null-point remains associated with the contrast-conserving (multiplicative) plaid (see Appendix). The Fourier components hypothesis thus predicts that the duty cycle manipulation should have little effect on coherency and that the multiplicative condition will again be associated with maximal motion non-coherency. The surface segmentation hypothesis offers markedly different predictions: as observed in Expt I, the contrast-reducing (additive) condition should be seen transparent and hence yield a high level of motion non-coherency. Conversely, the F/B reversal of what was previously seen as contrast-reduced yields a contrast-enhanced transparent configuration (bottom) that should yield reduced perceptual motion coherency. The multiplicative case, standing as the one luminance configuration that is consistent with transparency under both F/B interpretations, should exhibit approximately the same level of non-coherency for both duty cycle conditions.

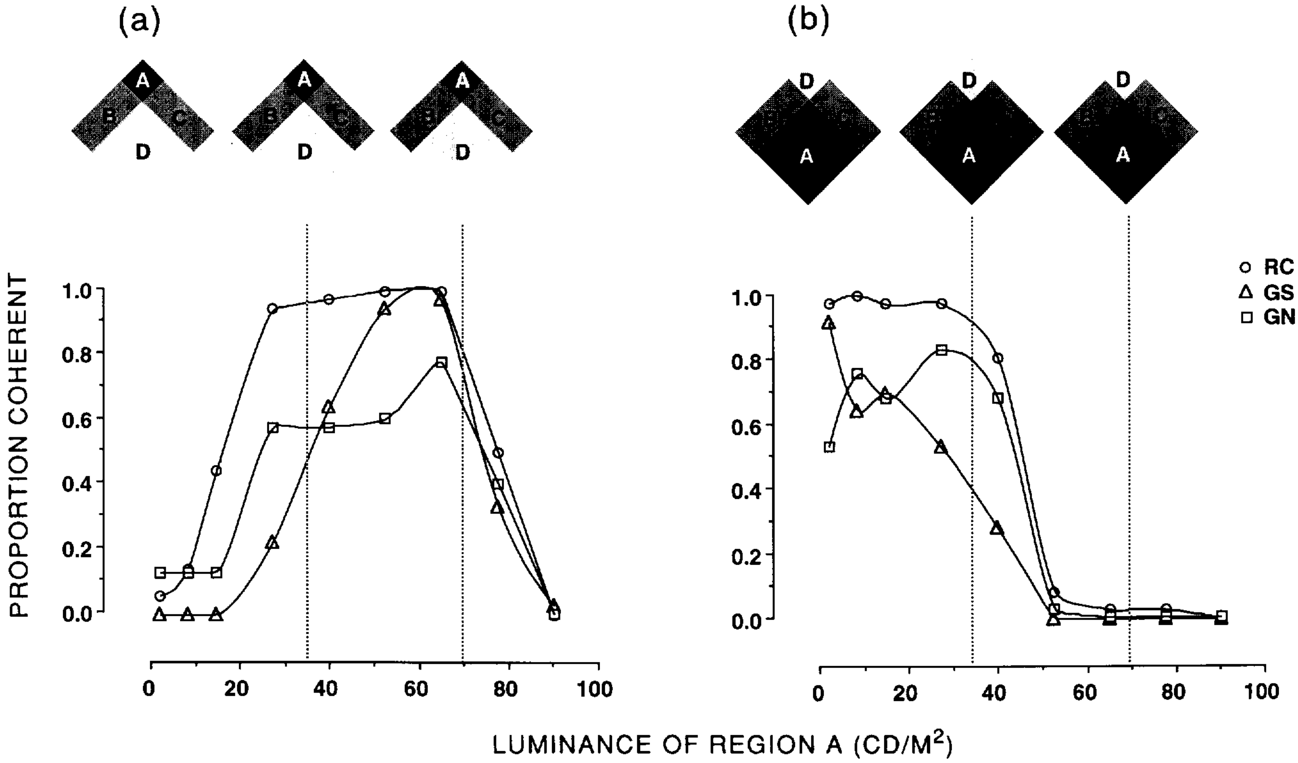

FIGURE 9. The conjoint effects of F/B interpretation and luminance configuration on motion coherency judgments for the moving plaid patterns used in Expt II (see Figs 7 and 8). As for the static transparency judgments (Expt I), two independent variables were manipulated: (i) the luminance of the region designated A, which is represented along the abscissa for each graph; (ii) the positional assignment of luminance-varying region A. For the graph in (a), region A corresponds to foreground; for the graph in (b), region A is background. All other stimulus properties, including the luminances of regions B, C, and D, were constant. The approximate luminance/size configuration for a subset of the different stimulus conditions is illustrated by the icons above each graph. Following each brief (1.33 sec) presentation of plaid motion, subjects were required to report judgments of perceptual motion coherence. The range of region A luminances that elicits a percept of non-coherent motion differs for the *A-as-foreground* (a) vs *A-as-background* (b) assignment and this range co-varies with physical compatibility with transparency. Note that, for a given region A luminance, the directional distribution of Fourier components is nearly unchanged by the F/B reversal (see Fig. 8). Data are shown for three subjects. Each data point is based on 40 trials. See text and legend to Fig. 6 for details.

turn, indicated the strength and directional distribution of Fourier components associated with each moving plaid pattern. Inevitably, small shifts in spectral content accompany the duty cycle manipulations used to alter F/B. Nonetheless, the LC for which Fourier energy moving in the pattern direction is nulled is independent of duty cycle (an analytical description of why this is true is provided in the Appendix). The Fourier components hypothesis thus predicts that our duty cycle manipulations will have little or no effect on the range of luminance values that render a percept of motion non-coherency. Conversely, the segmentation hypothesis predicts that duty cycle manipulations, which distinctly alter F/B interpretation and, hence, perceptual transparency, should determine which luminance values elicit motion non-coherency. These luminance values should conform to those established in Expt I.

Method

Visual stimuli and psychophysical procedure. Plaid stimuli were constructed as depicted in Fig. 7. The viewing aperture had a diameter of 13 deg. Both component gratings were of the same spatial frequency (0.31 c/deg) and were moved at an angle of 135 deg relative to one another at speeds of 2.70 deg/sec. Resultant pattern speed was 7.06 deg/sec. Pattern direction was either up or down and varied on a random schedule.

Two stimulus properties were manipulated as independent variables: the luminance of region A and the duty cycle of the component gratings. The luminance of region A took those values indicated in the General Method (same as for Expt I) and was varied from trial to trial on a random schedule. The luminances of all other subregions were held constant at the indicated values. Duty cycle was defined as: $B_{width} / (B_{width} + D_{width})$, where B and D correspond to the identified plaid sub-regions of Fig. 7 (B and C are interchangeable in this definition and in all references that follow, as both component gratings were identical and manipulated in parallel). Defined in this manner, the size of the luminance-varying sub-region (A) is proportional to duty cycle. Two different duty cycles were used in Expt II (0.2 and 0.8) and this parameter was held constant in separate blocks of trials. Other stimulus and procedural details, including the absolute luminances of regions A, B, C, and D, are provided above under General Method. Following each stimulus presentation, subjects were required to report

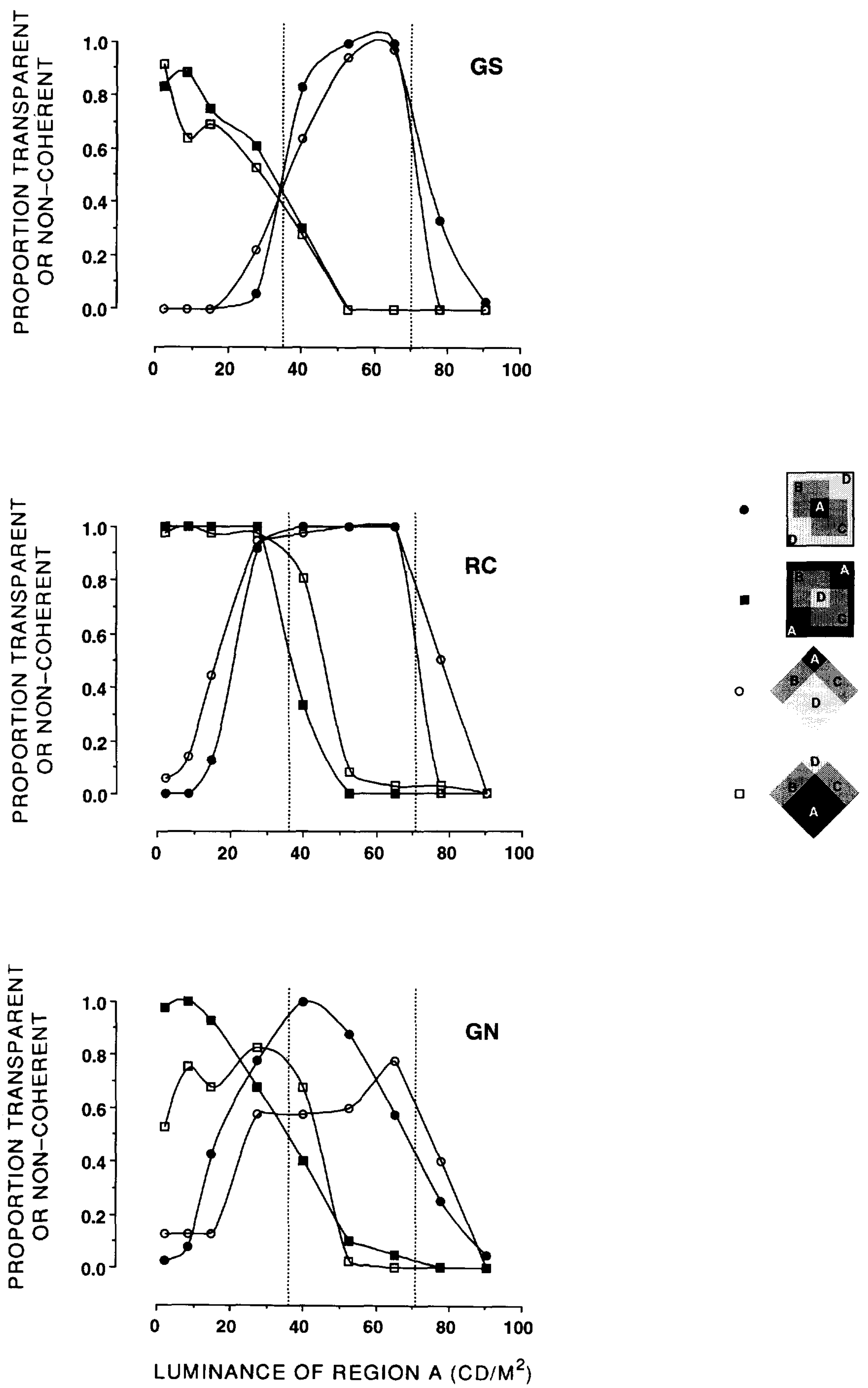

FIGURE 10. Luminance configuration and F/B conditions for transparency co-vary with those for motion coherency. The data from Expts I (Fig. 6) and II (Fig. 9) have been re-plotted on the same graphs to facilitate intra-subject comparison of the two types of judgments. For subjects GS and RC the two judgments showed a remarkable co-variance.

whether the two component gratings were seen to move coherently or non-coherently.

Results

All subjects reported a percept of motion coherency over a limited range of luminance values for each of the duty cycle conditions (Fig. 9). As was the case for the transparency judgments made in Expt I, these values were found to be similar across subjects and changed dramatically as a function of the duty cycle (F/B) manipulation.

Results obtained using the 0.2 duty cycle condition are shown in Fig. 9(a). Based upon previous evidence for the role of relative size in F/B interpretation (e.g. Koffka, 1935; Petter, 1956), this condition was expected to promote a percept in which region A was perceived as foreground and D as background. [Although the size-induced F/B assignment was not independently assessed in Expt II (in contrast to Expt III), the informal reports of our subjects confirmed this interpretation.] Under these F/B conditions, the luminance of region A can only be physically compatible with transparency/occlusion if its relationship to B is consistent with conservation or attenuation of the background contrast (expressed as C / D). Thus "physically transparent" values of A are bounded by $A / B = C / D$ [contrast conservation or "multiplicative transparency"; indicated by the left-most dotted line in Fig. 9(a)] and $A / B = 1.0$ [contrast elimination or opaque occlusion; indicated by the right-most dotted line in Fig. 9(a)]. For this 0.2 duty cycle condition, we found that region A luminance values eliciting maximum motion non-coherency were those consistent with physical transparency, as constrained by the size cue for F/B interpretation. This result was expected, as it constitutes a replication of the Stoner *et al.* (1990) experiment (although, see below).

Results obtained using the 0.8 duty cycle condition are shown in Fig. 9(b). In contrast to the effects of the 0.2 duty cycle, this condition was expected to promote a percept in which region A was perceived as background and D as foreground—an expectation consistent with informal subject reports. Following logic similar to that offered above, foreground region D (fixed luminance) can only be physically consistent with transparency/occlusion if its relationship to B is consistent with conservation or attenuation of the background contrast (now expressed as C / A). Thus "physically transparent" values of A—now background—are upper-bounded by $D / B = C / A$ [contrast conservation or multiplicative transparency; indicated by the left-most dotted line in Fig. 9(b)] and lower-bounded by $D = (B + C) - A$ [additive transparency; approx. $A = 0$ in Fig. 9(b)]. [The lower bound is given by the common-sense physical constraint that optical transmittance cannot exceed 100%. Hence, the luminance of the foreground image sub-region (D) cannot exceed the combined reflectances of the superimposed surfaces that contribute to it $(B + C)$] Notably, the physical transparency range imposed by this F/B condition is non-overlapping with that for the converse (0.2 duty cycle) F/B condition. As predicted by the image segmentation hypothesis, we found that the region A luminance values eliciting maximum motion non-coherency were those consistent with physical transparency as constrained by this size cue for F/B interpretation. These values were consistently and markedly *dimmer* than those that elicited maximum motion non-coherency for the converse F/B condition [cf Fig. 9(b) with Fig. 9(a)]—a result that is at variance with the Fourier components hypothesis.

Additional cause to reject the Fourier components hypothesis can be found in the precise positions and shapes of the psychometric functions. The specific assumption of a *logarithmic* compressive non-linearity, which has been incorporated into the Fourier components hypothesis (Noest & van den Berg, 1993; Wilson & Kim, 1994a, b), leads to the prediction that the *multiplicative* transparent condition should elicit maximal non-coherency. Contrary to this prediction, we observed maximal motion non-coherence in the 0.2 duty cycle condition [Fig. 9(a)] when LC deviated substantially from multiplicative transparency, and was roughly centered within the bounds of the transparency zone [Fig. 9(a)]. Interestingly, our previous study (Stoner *et al.*, 1990), which employed a slightly larger duty cycle (0.286), yielded luminance values for maximal non-coherency that were closer to multiplicative transparency. A tentative explanation for this difference can be found in the fact that the smaller duty cycle used in the present experiment is more effective in biasing F/B assignment, as is consistent with the greater A : D size asymmetry. The slight shift in the psychophysical function is to be expected if one grants that the percept of motion coherence associated with the 0.286 duty cycle was influenced to a greater extent by the alternative F/B interpretation.

In summary, Expt II revealed a striking interaction between plaid LC and duty cycle, which presents decisive testimony against explanations based on simple analyses of the strength and directional distribution of Fourier components. The results are entirely consistent with the image segmentation hypothesis. The latter interpretation is further supported by similarity between the shapes of the psychophysical curves obtained from individual subjects for the transparency (Expt I) and motion coherency (Expt II) judgments. These curves are replotted in Fig. 10 to facilitate intra-subject comparisons. This similarity is consistent with a common neural substrate subserving both types of perceptual judgments.

EXPERIMENT III: FOREGROUND/BACKGROUND INTERPRETATION AND MOTION COHERENCY: YOKED METASTABILITY

The results of Expt II preclude explanations for motion coherency founded on local non-linear image filtering and simple characteristics of spatio-temporal Fourier energy. Nonetheless, it is, in principle, possible to construct a model of this genre that relies upon somewhat

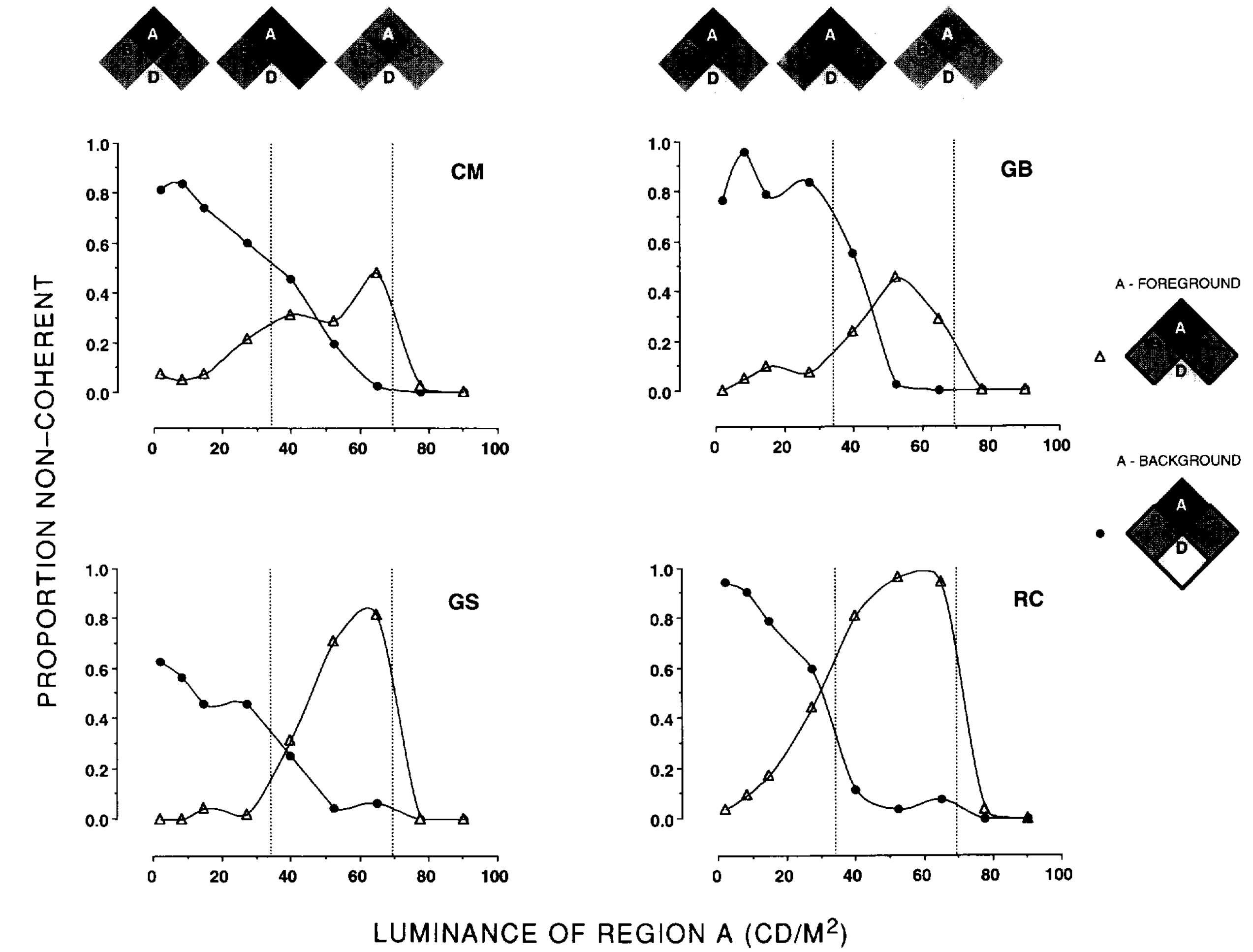

FIGURE 11. The conjoint effects of *metastable* F/B interpretation and luminance configuration on motion coherency judgments for the moving plaid patterns used in Expt III. In contrast to Expts I and II, in which F/B interpretation was manipulated via the relative size of image regions A and D, no size cue was presented in Expt III. Each stimulus thus supported either F/B interpretation and either could be seen—i.e. F/B interpretation was metastable. The only independent variable was the luminance of region A, which is represented along the abscissa for each graph. All other stimulus properties, including the luminances of regions B, C, and D, were constant and identical to those used in Expt II. The approximate luminance configuration for a subset of the different stimulus conditions is illustrated by the icons above the graphs. Following each brief (1.33 sec) presentation of plaid motion, subjects were required to report: (i) whether region A was perceived as foreground or background; and (ii) judgment of perceptual motion coherence. Coherence judgments were sorted and plotted by reported F/B interpretation: Coherence judgments for *A-as-foreground* trials are plotted with △, and those for *A-as-background* trials are plotted with •. Each graph illustrates data obtained from a single subject. As seen in Expt II, the range of region A luminances that elicits a percept of non-coherent motion differs for *A-as-foreground* vs *A-as-background*, and this range co-varies with attending physical compatibility with transparency. Notably, *for a given region A luminance, the stimulus is physically identical for the two F/B interpretations*, as is, of course, the directional distribution of Fourier components. Each data point is based on 40 or 42 trials. See text and legend to Fig. 8 for details.

more complex image information—such as conjunction of areal size and Fourier energy—which could account for these new results. While one might again point to the limited generality of such *ad hoc* explanations, their soundest dismissal can be sought in a disassociation between perceptual motion coherence and retinal image properties. It is well known from the fact of perceptual metastability that subjective assignment of foreground and background can be divorced from the specific properties of the retinal stimulus—the vase-face illusion (Rubin, 1921) is a classic example. Perceptual metastability can be similarly exploited in a further evaluation of the image segmentation hypothesis: does perceptual motion coherence co-vary with segmentation (F/B) fluctuations *in the absence of any changes in the retinal stimulus*? This question motivated Expt III.

A plaid possessing a symmetric (0.5) duty cycle offers no size cue for F/B interpretation. Not surprisingly, either region A or D (Fig. 7) can be seen as foreground, i.e. F/B assignment is metastable under these conditions. Positive support for the image segmentation hypothesis could come from a demonstration of tandem perceptual metastability of F/B assignment and motion coherence, in the face of constant retinal stimulation. To be specific, the defining prediction of this hypothesis is that F/B interpretation should determine which LCs elicit motion coherency judgments. To test this prediction, we allowed subjects to view symmetric (0.5) duty cycle plaids and required that they report both F/B interpretation and motion coherency for each trial. Because it is unconfounded by the changes (albeit minor) in the distribution of motion signals that accompany the stimulus manipulations of Expt II, and is not subject to any other explanations couched in terms of differing retinal content, this experiment constitutes the most definitive test for the role of surface segmentation mechanisms in motion coherency.

Method

Visual stimuli and psychophysical procedure. Preliminary testing indicated that perceptual reports of non-coherent motion are less common when viewing plaids with symmetric (0.5) duty cycles, as compared to asymmetric (0.2 or 0.8) duty cycles. To increase the likelihood of the non-coherent percept (and thereby elude potential ceiling effects) we increased the angular difference between component directions (to 157 deg) relative to that used in Expt II (135 deg), as this parameter is known to influence the probability of coherence (Adelson & Movshon, 1982). The pattern speed of 7.06 deg/sec was the same as for Expt II, whereas component speeds were decreased (to 1.41 deg/sec from 2.70 deg/sec), as a consequence of the change in component angle. All other stimulus parameters were identical to those used in Expt II and as defined in the General Method. The manner of stimulus presentation and psychophysical procedure were also unchanged, with one exception: in addition to the requirement for motion coherency judgments, subjects were instructed to report F/B interpretation. Specifically, they were asked to report whether region A (Fig. 7) appeared as the intersection of two gratings or as the background over which the two gratings moved. Subjects were instructed to report both judgments on each trial. Responses were recorded at trial termination via appropriate key-presses.

Results

Psychophysical judgments of motion coherence obtained from Expts I and II were grouped on the basis of *inferred* F/B interpretation, as predicted from figural cues and validated by informal questioning of subjects after completion of those experiments. Because the judgments of motion coherence obtained from Expt III can be grouped directly in terms of each subject's F/B reports, this experiment offers a more explicit test of the role F/B interpretation plays in motion coherency.

The results obtained when region A was interpreted as foreground are plotted separately for four subjects in Fig. 11 (△). The luminance of region *A-as-foreground* can only be physically compatible with transparency/occlusion if its relationship to B is consistent with conservation or attenuation of the background contrast (C / D). As in Expt II (and for reasons detailed in the Appendix), "physically transparent" values of *A-as-foreground* are bounded by $A / B = C / D$ (indicated by the left-most dotted line in each graph of Fig. 11) and $A / B = 1.0$ (indicated by the right-most dotted line in each graph of Fig. 11). As predicted, we found that the region A luminance values eliciting maximum motion non-coherency were those consistent with physical transparency, as constrained by F/B interpretation.

The results obtained when region A was interpreted as background are also plotted separately for the same four subjects in Fig. 11 (•). By contrast to the *A-as-foreground* condition, the luminance of *A-as-background* can only be physically consistent with transparency/occlusion if the (fixed) luminance relationship of foreground region D to B is consistent with conservation or attenuation of the background contrast (C / A). Thus "physically transparent" values of *A-as-background* are upper-bounded by $D / B = C / A$ (indicated by the left-most dotted line in each graph of Fig. 11) and lower-bounded by $D = (B + C) - A$ (approximately $A = 0$ in Fig. 11). The physical transparency range imposed by this F/B interpretation is non-overlapping with that for the converse interpretation. Once again, as predicted by the image segmentation hypothesis, we found that the region A luminance values eliciting maximum motion non-coherency were those consistent with physical transparency as constrained by F/B assignment.

Because the retinal stimulus is invariant across different F/B interpretations for a given LC, these data are clearly inconsistent with simple Fourier energy models of the type that have been proposed to account for transparency/motion coherence phenomena (e.g. Noest & van den Berg, 1993; see also Wilson & Kim, 1994a, b). These results, moreover, rule out the possibility that such models might be salvaged by simple modifications, such as, for example, the incorporation of particular spatial filters. Moreover, the coincident metastability of F/B interpretation and non-coherency judgments suggests that the neural mechanisms supporting these two types of perceptual representations communicate with one another, thereby ensuring consistency in the representation of depth-ordering and visual motion.

GENERAL DISCUSSION

Our data confirm that F/B interpretation is directly linked to perceptual motion coherency, lending key support to a growing body of evidence for the role of depth-ordering in visual motion signal integration. By contrast, our results offer no support for the Fourier-components hypothesis; indeed, they exclude the possibility that it can account for perceptual motion coherence under the conditions of our experiments. These experiments thus refine our understanding of the mechanisms responsible for motion signal integration, and they reinforce our ideas about the functional utility of this mechanism. In the remainder of this Discussion, we will address the relevance of these findings to: (i) other studies of a similar nature; (ii) the underlying neural mechanism and the domain of environmental phenomena over which it operates; (iii) non-Fourier motion; and (iv) the neuronal substrate for communication between depth-ordering and motion information.

Relationship to other studies

The viability of the Fourier components hypothesis depends upon the assumption that retinal image intensities are subject to a spatially localized and fixed logarithmic compression prior to motion detection. Mulligan (1993) conducted psychophysical experiments designed to investigate this possibility. Employing overlapping fields of randomly positioned dots that were moved in opposite directions, Mulligan found that maximal dot field segregation occurred when dot luminance was combined according to near-additive rules. At first glance, these findings appear at odds with the results of Stoner *et al.* (1990), in which it was reported that perceptual non-coherence (analogous to dot field segregation in Mulligan's experiments) is nearly non-existent for additive plaids, and highly likely when multiplicative rules were used for luminance combination.

Our latest analysis provides an explanation for this apparent discrepancy: maximal segmentation is not associated with any *fixed* luminance combination rule (be it additive or multiplicative). Rather, the rule that renders maximal segregation is contingent upon other cues for F/B assignment. By definition, random-dot stimuli—unlike plaid patterns—lack the spatial structure necessary for F/B interpretation. It follows from the arguments made herein that random-dot patterns also lack the ability to support luminance-based transparency. Not surprisingly, static versions of these stimuli failed to generate a percept of transparency using any luminance combination rule (Mulligan, 1993). Thus in the absence of appropriate segmentation cues, the motion system operates on a linear representation of its retinal input. In the presence of such cues, as in the experiments of Stoner *et al.* (1990), non-linear mechanisms prevail.

The issue of non-linear pre-processing comes up again in a recent study by Noest and van den Berg (1993), in which, like the present study, the influence of duty cycle on motion coherence was examined. There are two noteworthy aspects to their results. First, in contrast to Mulligan (1993), they found that the *multiplicative* superimposition condition resulted in maximal non-coherent motion. Second, in seeming contradiction to our present findings, they reported that duty cycle manipulations *did not* alter the relationship between LC and perceived motion of plaid patterns. Together, these observations were interpreted as evidence for the existence of a *fixed* logarithmic compression of image intensity prior to motion processing.

We can account for the discrepancy between the results of Noest and van den Berg and those of Mulligan (multiplicative vs additive rules) using the same logic applied above. The failure of Noest and van den Berg to find an effect of duty cycle is more puzzling. In view of the robustness of our results, we must conclude that procedural and stimulus differences account for this discrepancy. Subjects in the Noest and van den Berg experiments were instructed to report their ability to *detect* motion in either the component or pattern direction. By contrast, our subjects were required to judge *coherence*. The latter is arguably a more direct measure of segmentation. Perhaps more importantly, the plaid stimuli used by Noest and van den Berg differed from our own by superimposition of a dynamic noise pattern. As Noest and van den Berg offer no assurance that the ability of duty-cycle to bias F/B interpretation survives these degraded stimulus conditions, evaluation of their results must remain tentative.

Finally, Trueswell and Hayhoe (1993) also attempted to evaluate the relative merits of the Fourier components and segmentation hypotheses. Adopting a strategy conceptually similar to our own, these investigators employed two different cues for depth-ordering and examined the consequences of cue agreement vs conflict for perceptual motion coherence. The first cue was LC. In using this cue, Trueswell and Hayhoe took advantage of the fact that certain transparent LCs [i.e. those in which the luminance of region B differs from that of C, unlike the symmetric (B = C) conditions of our experiments (Fig. 3)] render an apparent depth ordering of the superimposed gratings. The second cue was horizontal binocular disparity, which was introduced as a means to manipulate stereoscopic depth ordering of the two gratings. The disparity cue was adjusted such that it either agreed or conflicted with the luminance-based depth-ordering cue. A percept of non-coherent motion was most common when there was agreement between the two cues. In other words, maximal motion non-coherency was not associated with a fixed intersection luminance. These results are incompatible with simple mechanisms based solely upon the directional distribution of moving Fourier components. They do, however, stand as further evidence for cooperativity between the visual motion system and surface segmentation mechanisms.

Why transparency? Ecological relevance and potential neural mechanisms

Simple reflection on the task facing the visual system makes it clear that neural mechanisms must exist for the appropriate generation of multi-valued or "segmented" representations. Such mechanisms, we assert, must avail themselves of retinal image segmentation cues that reflect the spatial arrangement of surfaces in natural visual scenes. Given the seeming rarity of transparent surfaces in the natural world, however, one might question the notion that luminance-based cues for perceptual transparency play a critical role in the segmentation of moving visual stimuli. In view of the accepted status of transparency for static images, however, we see no reason to view the suggested importance of transparency for the interpretation of dynamic visual stimuli with increased skepticism. Nonetheless, the question remains: Why should the visual system exhibit sensitivity to luminance cues for transparency? Following von Helmholtz's (1860/1924) lead, Noest and van den Berg (1993) suggested that this sensitivity may simply reflect the workings of a mechanism that normally operates to discount the illuminant. This seems a sensible proposition, given the ubiquity of shadows in natural images, and the fact that shadows "mimic" the LC of multiplicative transparency. The findings presented here, however, demonstrate the importance of non-multiplicative transparency and thus argue against shadows being the exclusive or even the primary realm of application.

Another plausible explanation arises from the fact that many surfaces, such as wire screens with very fine meshes, appear transparent simply due to a failure to resolve the individual opaque occluders that make up those surfaces. Indeed, it is generally true that appropriate low-pass filtering of the image of any object (such as a bush, a tree or a spider's web) composed of many small opaque occluders, can render that image consistent with a single large transparent foreground object. It may be that the process responsible for inferring depth-ordering from such images operates on a scale coarser than that available perceptually. The hypothesis that opaque occlusion is the primary domain of the process subserving perceptual transparency receives strong support from the findings of Trueswell and Hayhoe (1993).

Potential contribution of non-Fourier motion mechanisms

The term "non-Fourier motion" has been used to refer to a broad class of stimuli that are (by definition) invisible to motion detection mechanisms that sense spatio-temporal Fourier energy, i.e. "Fourier motion" (Chubb & Sperling, 1988). The results of recent psychophysical (Ledgeway & Smith, 1994) and neurophysiological experiments (Zhou & Baker, 1993) suggest that Fourier and non-Fourier motions are initially processed by independent pathways. Conventional plaid patterns (including those configured to mimic transparency) would be expected to stimulate both Fourier and non-Fourier pathways (Wilson *et al.*, 1992). In view of these considerations, a question naturally arises as to the potential contribution of a non-Fourier pathway to the perceptual coherence of moving plaids. Wilson and Kim (1993) have proposed a model that combines Fourier with non-Fourier signals at the motion integration stage. That mechanism is *not* sufficient to account for the results presented herein: like their Fourier counterparts (see Fig. 8 and Appendix), the magnitude of non-Fourier pattern-direction components is minimal for plaid stimuli configured to mimic multiplicative transparency (if the assumption of a logarithmic transformation of image intensity is incorporated). As for Fourier signals, this remains true regardless of duty cycle; hence, regardless of F/B interpretation.

Our results, as well as our physically grounded observations concerning transparent superimposition, suggest a novel role for the non-Fourier pathway. In particular, we propose that a primary function of this pathway is to provide the visual system with depth-ordering information. Two observations support this hypothesis. The first is recognition of a physical parallel between a sub-type of non-Fourier image variation and the consequences of transparent occlusion in natural images: the former is defined by luminance contrast modulation (Chubb & Sperling, 1988) and the latter consists of the same. Indeed, the opacity of a foreground surface is directly encoded by the depth of the contrast modulation within the non-Fourier elements (see Appendix). The second supportive observation is empirical: motion of image features defined by contrast modulation yield an impression of a foreground surface of variable opacity traveling over a static background. This promotion of depth ordering by non-Fourier motion has recently been shown to be extremely potent (Stoner & Albright, 1995) as revealed by its ability to override depth ordering supported by binocular disparity cues. By providing information about the depth ordering of dynamic scene elements, non-Fourier motion thus provides information that could be utilized in the formation of segmented representations (Stoner & Albright, 1992b).

Whence comes image segmentation? Computational considerations and neuronal mechanisms

Whether or not non-Fourier mechanisms are proved culpable, a key question concerns the means by which depth-ordering mechanisms influence motion coherency. We will consider three related possibilities. Firstly, depth-ordering information could influence motion coherency by allowing identification of image features resulting from occlusion. The retinal consequence of occlusion is the production of extrinsic image components, the motions of which must be "ignored" for veridical scene interpretation. One plausible strategy is to suppress the processing of extrinsic features. This basic approach can be found in a recent model that employs "selection units" to distinguish reliable (i.e. intrinsic) from unreliable (i.e. extrinsic) motion signals (Nowlan & Sejnowski, 1994, 1995). These selection units have been tentatively identified with pattern-type neurons in area MT.

Rather than suppressing the processing of extrinsic features, another possibility is that depth-ordering cues foster a reconstructed neural representation of occluded background scene elements. Support for this idea comes from psychophysical experiments by Shimojo and Nakayama (1990), who found that depth-ordering cues promote a representation of the occluded surface that impacts motion perception in a manner indistinguishable from non-occluded surfaces. By correctly reinterpreting extrinsic image features in terms of both foreground and background motions, the influence of these spurious image motions might be removed.

A third potential role for occlusion in motion processing is the "binding" of motion signals that arise from a common surface. A controversial mechanism that has been proposed as a means to achieve such binding is the control of phase relationships (i.e. relative synchrony) between neuronal firing patterns (e.g. Gray *et al.*, 1989). According to this proposal, neurons encoding image regions corresponding to a single object should fire synchronously. The likelihood of a single-valued representation of motion at the integration stage (MT pattern-type neurons) would thus depend upon the relative synchrony of ascending inputs from the motion detection stage. The latter would be regulated by mechanisms sensitive to image segmentation cues. Much more work is needed to evaluate this feature binding hypothesis; the stimulus manipulations described herein would seem to offer an ideal experimental paradigm.

Another fundamental issue on which speculation is warranted is the neuroanatomical locus (loci) of the image segmentation mechanism(s). Given knowledge of the types of image factors that affect segmentation, several lines of evidence implicate area V2. For example, von der Heydt *et al.* (1984) have found that a subset of area V2 neurons respond to "illusory contours" that arise from occlusion. V2 neurons, in addition, are sensitive to binocular disparity (DeYoe & Van Essen, 1985)—another stimulus parameter with an obvious connection to depth ordering and a strong influence on motion coherency. Even more intriguing is a recent report asserting that many V2 neurons are selective for the "sidedness" of an edge (i.e. which side is foreground and which background) (Peterhans & von der Heydt, 1992). Evidence indicates that neurons exhibiting illusory contour and/or binocular disparity sensitivity are found most frequently in the thick cytochrome-oxidase stripes of area V2 (Peterhans & von der Heydt, 1993). This functional compartment is known to provide a substantial ascending projection to area MT. This projection may thus be the source of the depth-ordering information implicated in our studies of the neural correlates of motion coherency (Stoner & Albright, 1992a; Duncan *et al.*, 1995).

Why are there multiple cues for perceptual motion coherence?

It is clear that the stimulus parameters focused on in this report are not the sole determinants of motion coherency; abundant data document the influence of a wide variety of factors. Elsewhere we have argued that the influence of a subset of these parameters (duty cycle, spatial frequency, luminance contrast) may reflect the operation of mechanisms sensitive to occlusion (see Stoner & Albright, 1993, unpublished observations). Velocity variation is, however, a potent determinant of motion coherency (Adelson & Movshon, 1982) and image segmentation (e.g. Braddick, 1993) that possesses no straightforward relationship to occlusion-based surface segmentation. Based on the observation that a plaid configured to mimic transparent occlusion may cohere if its components move in similar directions, Kim and Wilson (1993) state that, "This angular dependence also explains recent data previously thought to be based on a visual computation of multiplicative transparency". Contrary to this conclusion, evidence indicates that occlusion- and velocity-based determinants of motion coherence operate independently. Thus, while decreasing the angular difference between components of a plaid increases overall levels of coherence, the LC that elicits maximal non-coherency remains constant (Stoner & Albright, 1995). These observations argue that, although the absolute strength of perceptual coherency may vary, the ordinal effects we report are not peculiar to (nor can they be explained by) a specific choice of component angles.

We submit that interactions between velocity and occlusion-based surface segmentation cues are best understood with reference to the composition of natural scenes. Natural images typically contain moving components of both intrinsic and extrinsic varieties. We have suggested that occlusion cues may influence motion perception by permitting reconstruction of the occluded background, thereby nulling extrinsic image components. From this proposal, it follows that the ability of occlusion cues to bias motion perception should depend upon the relative distribution and strength of extrinsic vs intrinsic image components. For example, due to the relatively broad directional tuning of neurons sensitive to visual motion, plaids possessing components that move in similar directions would be expected to produce a nearly unimodal pattern of neuronal activation. Wilson and Kim's observation that these type of stimuli may cohere despite the presence of surface segmentation cues is perhaps, therefore, not very surprising. Natural scenes dense with occlusion (e.g. a predator moving through foliage), however, would not be expected to give rise to uniformly unambiguous velocity distributions. For many regions of such an image, the presence or absence of extrinsic image components may determine whether the associated pattern of neuronal activation is unimodal or bimodal. Under these circumstances, the contribution of surface segmentation mechanisms may be very critical indeed. It should be kept in mind that it is, after all, for such natural scenes (not plaid patterns) that the visual motion system was "optimized".

REFERENCES

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

Albright, T. D. & Stoner, G. R. (1995). Visual motion perception. *Proceedings of the National Academy of Science, 92*, 2433–2440.

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

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends in Neuroscience, 16*, 263–268.

Chubb, C., & Sperling, G. (1988). Drift-balanced stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5*, 1986–2006.

De Valois, K. K., De Valois, R. L. & Yund, E. W. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *Journal of Physiology, London, 291*, 483–505.

DeYoe, E. A. & Van Essen, D. C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature, 317*, 58–61.

Duncan, R. O., Stoner, G. R. & Albright, T. D. (1995). Gating of perceptual motion coherence by foreground/background interpretation: Neural correlates in area MT. Society of Neuroscience Abstracts. In press.

Gray, C. M., Koenig, P., Engel, A. K. & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338*, 334–337.

von Helmholtz, H. (1860/1924). *Physiological optics* (Vol. 3). English translation by Southall, J. P. C. for the Optical Society of America from the 3rd German edition of 1909, *Handbuch der Physiologischen Optik*. Hamburg: Voss.

von der Heydt, R., Peterhans, E. & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.

Kersten, D., Bülthoff, H. H., Schwartz, B. L. & Kurtz, K. J. (1992). Interaction between transparency and structure from motion. *Neural Computation*, *4*, 573–589.

Kim, J. & Wilson, H. R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research*, *33*, 2479–2489.

Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge & Kegan Paul.

Kooi, F. L., De Valois, K. K. & Switkes, E. (1992). Higher order factors influencing the perception of sliding and coherence of a plaid. *Perception*, *21*, 583–598.

Ledgeway, T. & Smith, A. T. (1994). Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Research*, *34*, 2727–2740.

MacLeod, D. I. A. (1978). Visual sensitivity. *Annual Review of Psychology*, *29*, 613–645.

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

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 91–95.

Movshon, J. A., Adelson, E. A., Gizzi, M. & Newsome, W. T. (1985). The analysis of moving visual patterns. In Chagas, C., Gattass, R. & Gross, C. G. (Eds), *Study group on pattern recognition mechanisms* (pp. 117–151). Vatican City: Pontifica Academia Scientiarum.

Mulligan, J. B. (1993). Non-linear combination rules and the perception of visual motion transparency. *Vision Research*, *33*, 2021–2030.

Noest, A. J. & van den Berg, A. V. (1993). The role of early mechanisms in motion transparency and coherence. *Spatial Vision*, *7*, 125–147.

Nowlan, S. J. & Sejnowski, T. J. (1994). Filter selection model for motion segmentation and velocity integration. *Journal of the Optical Society of America*, *11*, 3177–3200.

Nowlan, S. J. & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *Journal of Neuroscience*, *15*, 1195–1214.

Peterhans, E. & von der Heydt, R. (1992). Selectivity for figure–ground direction at occluding contours in monkey area V2. *Society for Neuroscience Abstracts*, *18*, 1275.

Peterhans, E. & von der Heydt, R. (1993). Functional organization of area V2 in the alert macaque. *European Journal of Neuroscience*, *5*, 509–524.

Petter, G. (1956). Nuove ricerche sperimentali sulla totalizzazione percettiva. *Rivista di psicologia*, *50*, 213–227.

Plummer, D. J. & Ramachandran, V. S. (1993). Perception of transparency in stationary and moving images. *Spatial Vision*, *7*, 113–123.

Rodman, H. R. & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, *75*, 53–64.

Rubin, E. (1921). *Visuell Wahrgenommene Figuren*. Glydendalske.

Shimojo, S. & Nakayama, K. (1990). Amodal representation of occluded surfaces: Role of invisible stimuli in apparent motion correspondence. *Perception*, *19*, 285–299.

Shimojo, S., Silverman, G. H. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, *29*, 619–626.

Stoner, G. R. & Albright, T. D. (1992a). Neural correlates of perceptual motion coherence. *Nature*, *358*, 412–414.

Stoner, G. R. & Albright, T. D. (1992b). The influence of foreground/background assignment on transparency and motion coherence in plaid patterns. *Investigative Ophthalmology and Visual Science (Suppl.)*, *33*, 1050.

Stoner, G. R. & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, *5*, 129–149.

Stoner, G. R. & Albright, T. D. (1994). Visual motion integration: A neurophysiological and psychophysical perspective. In Smith, A. T. & Snowden, R. J. (Eds), *Visual detection of motion*. London: Academic Press.

Stoner, G. R. & Albright, T. D. (1995). Depth, dynamic occlusion and non-Fourier motion. *Society of Neuroscience Abstracts*. In press.

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

Trueswell, J. C. & Hayhoe, M. M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research*, *33*, 313–328.

Vallortigara, G. & Bressan, P. (1991). Occlusion and the perception of coherent motion. *Vision Research*, *31*, 1967–1978.

Wallach, H. (1935). Ueber visuell wahrgeommene bewegungsrichtung. *Psychologische Forschung*, *20*, 325–380.

Wilson, H. R., Ferrera, V. P. & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.

Wilson, H. R. & Kim, J. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research*, *33*, 2479–2489.

Wilson, H. R. & Kim, J. (1994a). Perceived motion in the vector sum direction. *Vision Research*, *34*, 1835–1842.

Wilson, H. R. & Kim, J. (1994b). Corrigendum. *Vision Research*, *34*, 973.

Zhou, Y. X. & Baker, N. K. (1993). A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science*, *261*, 98–101.

Acknowledgements—Supported by NIMH grant MH50706 and a Development Award from the McKnight Endowment Fund for Neuroscience. Gene Stoner was partially supported by a Research Fellowship from the McDonnell-Pew Center for Cognitive Neuroscience at San Diego. We thank G. Buracas, L. Croner, R. Duncan, T. Sejnowski, and L. Stefanacci for helpful discussion and comments on the manuscript. J. Costanza provided superb technical assistance.

APPENDIX

Fourier components in moving plaid patterns.

In most previous experiments utilizing plaid patterns (e.g., De Valois *et al.*, 1979; Adelson & Movshon, 1982; Movshon *et al.*, 1985; Rodman & Albright, 1989), visual stimuli were constructed by *additive* superimposition of the luminances of two component gratings. In contrast to this conventional method of construction, plaids were produced in the Stoner *et al.* (1990) study by simply increasing the luminance of region A (Fig. 3) beyond that which would result from additive superimposition. Many of the resultant plaids were found to be perceptually transparent. It is useful to regard such stimuli as being comprised of a conventional "additive" plaid pattern plus an additional lattice of "intersections." Borrowing the convenient formalization of Noest and van den Berg (1993), the two-dimensional spatial luminance profile can be described as:

$$\text{Plaid}\,(x,y) = \underbrace{[D + (B - D) * (G_0(x,y) + G_1(x,y))]}_{\textit{Conventional Plaid}} + \underbrace{[(A - 2B + D) * I(x,y)]}_{\textit{Intersections}} \quad \text{(A1)}$$

where $G_0(x,y), G_1(x,y), I(x,y) 0, 1$; $I(x,y) = G_0(x,y) * G_1(x,y)$; $G_0(x,y)$ and $G_1(x,y)$ are square-wave gratings; $I(x,y)$ defines the spatial positioning of the intersections. A, B ($C = B$), and D correspond to the regional luminance values indicated in Fig. 3.

It can be readily shown that a conventional plaid, created by linear superimposition, possesses only those Fourier components associated with the two oriented gratings (Fig. 4). For such plaids, the luminance of intersection region $A = 2B - D$, i.e. the quantity $A - 2B + D = 0$.

Thus the lattice of intersections [identified by the Intersection Term in Eqn (A1)] has an luminance amplitude of zero. The addition of a lattice of intersections with non-zero amplitudes introduces oriented Fourier components that move in the pattern direction. The strength of this potential motion signal increases with the quantity $A - 2B + D$ [Intersection Term, Eqn (A1)]. If motion coherency were strictly determined by the directional distribution of moving Fourier components, coherence should be minimal when $A = 2B - D$, as occurs for the additive or conventional plaid. The results of Stoner et *al.* (1990) show that not to be the case.

An explanation of motion coherency couched in terms of Fourier components might still be salvageable, however, if we hypothesize the existence of a local non-linear transformation of image intensity that occurs prior to motion detection. [Evidence suggests that this may be a viable hypothesis and that the form of the non-linearity may be logarithmic (MacLeod, 1978; MacLeod *et al.*, 1992).] Clearly, the values of transformed image intensities (A, B, and D) that are required to cancel the Intersection Term in Eqn (A1) are not those associated with additive superimposition of the component gratings. In particular, if the transformation were logarithmic, the real image must possess luminance relationships of the form $A = B^2 / D$ in order for the transformed contributions from the intersections to be zero (i.e. $A' = 2B' - D'$ only if $A = B^2 / D$). In other words, the intersection term of the logarithmically compressed plaid will be nulled only when the plaid stimulus is created by multiplicative superimposition of the component gratings. This change in the null-point can also be appreciated by examining the Fourier components associated with different intensities of image region A, *with and without* logarithmic transformation of intensity (Fig. 4). The shifting of the null-point to multiplicative plaids, imposed by a log compression of image intensities, leads the Fourier components hypothesis to predict that perceptual motion coherence should be minimal when viewing multiplicative plaids. This prediction appears generally consistent with the results of Stoner *et al.* (1990).

Non-Fourier components in moving plaid patterns. An image arising from foreground occlusion of background can be described as the sum of two components:

$$I(x,y) = \overset{[i]}{[M_B + (R_F - M_B) * F(x,y)]} + \overset{[ii]}{[(1 - F(x,y)) * B(x,y) * (R_B - M_B)]}; \quad \text{(A2)}$$

where terms $B(x,y)$ and $F(x,y)0,1$ and can be thought of as the "density", at a given two-dimensional spatial location, of the material comprising background and foreground surfaces. R_F and R_B correspond to the light reflected off of those materials. $F * R_F$ and $B * R_B$ hence describe the respective luminances of background and foreground surfaces viewed in isolation. M_B is the mean luminance of the background. The first component [i] thus describes the foreground with luminance amplitude $(R_F - M_B)$. The second component [ii] is the product of foreground opacity and background luminance variation (modulated around zero), and corresponds to the contrast modulation of the background by the foreground occluder. Evidence suggests that the visual system has specialized mechanisms for the extraction of contrast-modulated image components (Zhou & Baker, 1993). Since the transition from high to low luminance contrast modulation is a direct indicator of the location of foreground objects, that "non-Fourier" mechanism may be providing information about the depth ordering of scene elements.