Pergamon

0042-6989(93)E0041-5

Vision Res. Vol. 34, No. 16, pp. 2125-2135, 1994
Copyright © 1994 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0042-6989/94 $7.00 + 0.00

Perceived Surface Shape not Features Determines Correspondence Strength in Apparent Motion

ZIJIANG J. HE,* KEN NAKAYAMA*

Received 20 September 1993; in revised form 23 November 1993

Previous psychophysical studies have revealed that shape similarity can affect apparent motion correspondence. Such results however, do not specify the level of representation, at which shape similarity is defined. We sought to understand this question by using a 2 × 2 competitive apparent motion paradigm. We manipulated the binocular disparity of the motion stimuli (tokens) relative to adjacent squares to selectively change the internal surface representation of the tokens while keeping early filtered representation intact. When two sets of differently oriented tokens (45°, −45° bars) were used, there was a preference for seeing motion between tokens having the same orientation. However, such a motion bias was reduced when tokens became part of a large surface square seen either as amodally occluded in the background or as a transparent surface modally completed in front. Since shape differences at the early filtering level remain essentially intact (i.e. +45° vs −45°) our findings support the surface level hypothesis. Perceived surface shape rather than shape defined by early filters largely determines motion correspondence.

Apparent motion   Early filtering   Surface representation   Transparency   Binocular disparity

INTRODUCTION

Apparent motion is perceived when two stationary stimuli are presented sequentially. If multiple elements are presented at different times, the visual system has to solve the correspondence problem, viz. it must determine which two successive stimuli represent the same object over time (Anstis, 1980; Braddick, 1980; Ullman, 1979). This correspondence problem can be illustrated in Fig. 1(A) (Ramachandran & Anstis, 1983). At time T1, two tokens (squares) are displayed at the diagonal corners of an imaginary rectangle. At time T2, another pair of tokens is presented at the remaining corners. The visual system now faces the choice of matching the tokens at T1 with either their horizontal or vertical neighbors presented at T2, and each alternative in turn, will lead to a radically different perception of motion, of horizontal or vertical motion, respectively (Fig. 1).

Motion correspondence strength between motion stimuli is generally believed to depend on how each stimulus is spatially represented internally in the brain. For instance, it could be a representation of distance (visual angle) between the stimuli in successive frames: when there are several possible matches, a stimulus tends to correspond to the one located nearest to it (proximity rule) (Burt & Sperling, 1981; Ullman, 1979). Motion correspondence can also be influenced by the surface layout relationship between the motion stimuli in 3-D space. Recently, we have demonstrated that when the 3-D motion stimuli were placed on the same surface, their matching affinity became stronger (He & Nakayama, 1994).

Form similarity, the focus of this paper, has also been found to affect apparent motion correspondence strength (Green, 1986; Prazdny, 1986; Ramachandran, 1985, 1988). This stems from the idea that when the internal representations of two stimuli possess similar properties, a stronger motion correspondence between them will be observed. Such a result however, does not reveal the level of representation at which form similarity is determined. It could for example, result from the properties of an early cortical filtering stage, or of a later stage of surface representation, or beyond. In this paper, we will approach this problem by studying a motion display whose internal representations of motion tokens at the early filtering and surface representation stages will predict different motion directions, thus enabling a dissociation of the two.

Kanizsa (1979), following the earlier work of Michotte (1954), had classified surface completion phenomena into two distinct categories. Each is defined in terms of whether the surfaces are made visible and "complete" as

*Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, U.S.A. [*Email:* zh@wjh12.harvard.edu].

occluders in front of other surfaces (modal completion), or whether they are invisible yet are seen to complete as occluded surfaces behind other surfaces (amodal completion).

Modal completion is exemplified in the well known Kanizsa triangle. Even though there is no triangle, the contours of an illusory triangle "complete", occluding the disks. Modal completion can be strengthened with binocular stereopsis, as seen by fusing the Kanizsa triangle in Fig. 2. In this paper, we also deal with another type of modal completion where the "completion" of a surface is remarkably strong. Here an otherwise black surface patch is seen to be covered by a colored transparent surface [to see this look ahead to Fig. 7(B)] (Nakayama & Shimojo, 1990; Nakayama, Shimojo & Ramachadran, 1990).

Amodal surface completion is demonstrated in Fig. 3. After free fusing the top stereogram [Fig. 3(A)], the reader will see black L-shaped tokens in front of their adjacent white squares. But when viewed in the bottom stereogram [Fig. 3(B)], the L-shaped tokens are no longer seen as Ls, but are completed and perceived as black squares in back, occluded by their adjacent white squares. These stereograms demonstrate the ability of binocular disparity to alter perceived surface shape, while leaving the neural representation at the early filtering level essentially unaffected. Following this, our approach is to use binocular disparity to manipulate the surface completion phenomena in order to understand the differential roles of the early filtering and surface representation levels in visual information processing.

The following description further explains this approach. The black Ls will be used as motion tokens in a 2 × 2 competitive apparent motion paradigm, where each diagonal pair of motion tokens (black Ls) will be presented in two alternate frames. This will lead to an apparent motion perception of the tokens moving in either a horizontal or vertical direction. The preference for motion direction, horizontal vs vertical, depends on the relative motion correspondence strength in the two directions. If correspondence is stronger between the left and right tokens as compared to the top and bottom tokens, the perceived motion direction will be horizontal. Specifically, in the top stereogram of Fig. 3, where the motion tokens are seen in front, we would expect a horizontal motion bias. This is because the horizontal pairs of tokens, and not the vertical ones, have identical L-shaped representations at the early filtering level, as well as at the surface representation level. However, different neural representations at the two levels exist for the tokens in the bottom stereogram of Fig. 3. Due to amodal surface completion, the black L-shaped tokens are now internally represented as black squares, even though their internal representation at the early filtering level remains unchanged. In other words, the distinct orientation difference between the top and bottom pairs of motion tokens seen at the early filtering level disappears at the surface representation level. So, if early filtering level determines the apparent motion matching process, we would expect a similar horizontal motion bias in the back case as in the front case (early filtering hypothesis). But, if matching is determined at the surface representation level, we would predict less horizontal motion bias in the back case compared to the front case (surface hypothesis).

GENERAL METHODS AND PROCEDURES

The stereo motion stimuli were displayed on a TV monitor, which was driven by a Commodore (A2000) computer, and viewed through a pair of haploscopic prisms, at a viewing distance of 100 cm. A three dimensional 2 × 2 competitive apparent motion paradigm was used in the experiments (see Fig. 3 for example). All the motion tokens had the same binocular disparity. Each diagonal pair of tokens was presented alternately (for 300 msec, except in Experiment 3) in 6 frames (3 repetitions of each diagonal pair of stimuli) on each trial.

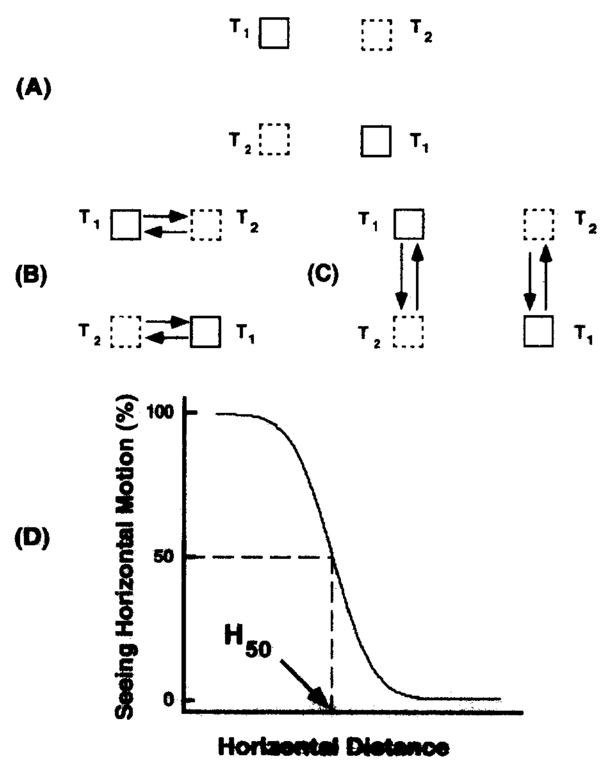

FIGURE 1. A bi-stable 2 × 2 apparent motion display where stationary target are presented alternately at T1 and T2 (A). At time T1, two tokens (squares) are displayed at the diagonal corners of an imaginary rectangle. At T2, another pair of token is presented at the remaining corners. The visual system now has to solve the correspondence problem, viz. it must determine which two successive stimuli represent the same object over time, i.e. matching tokens at T1 with either their horizontal or vertical neighbors presented at T2. Each alternative in turn will lead to a radically different perception of motion, either horizontal or vertical motion, respectively. One critical factor determining motion correspondence is the perceived distance between stimuli in successive frames. As illustrated in (B), when the vertical distances are kept the same, short horizontal distances will result in a horizontal match, and consequently will produce a horizontal motion; longer distances will lead to vertical motion (C). This "proximity" tendency for the motion token to match its nearest neighbor, can be summarized by a motion dominance function that denotes the percentage of seeing horizontal motion at each horizontal distance (D).

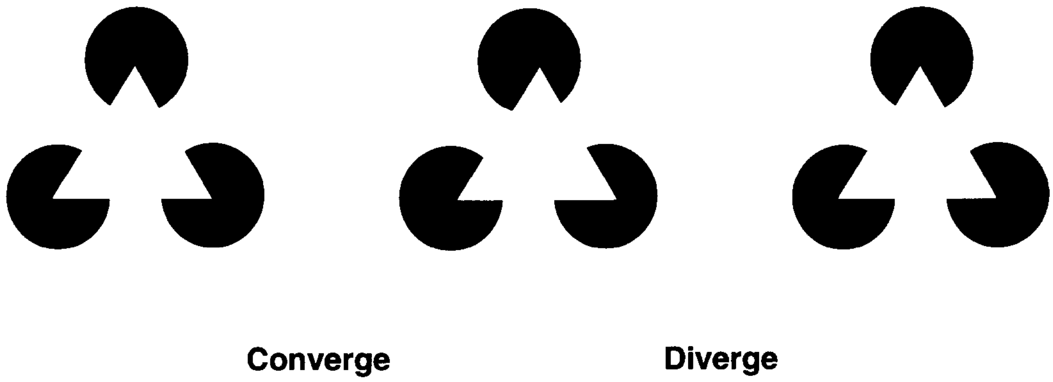

FIGURE 2. Illustration of the impact of binocular disparity on the perception of Kanizsa illusory triangle. After free fusing the two left Kanizsa illusory triangles convergently, or the two right ones divergently, the reader will perceive an illusory triangle modally completing in front of three partially occluded disks. Note that the triangle is perceived as more vivid after fused.

The perception of motion under these conditions is bi-stable and depends on whether a given token matches a horizontal or vertical neighbor. A small black cross, used as the fixation point, was inserted at the same fronto-parallel plane as the motion tokens. During the experiments the observer was instructed to gaze at the fixation point and to judge the apparent motion direction (vertical or horizontal) after the tokens had disappeared from the screen. Each test condition (i.e. one block) consisted of 96 trials. In each trial, the center-to-center vertical distance between the top and bottom pairs of tokens was held constant; the horizontal separation however, could assume one of six distances, scheduled in cyclical sequence, for a series of 6 trials the horizontal distances were incremented, followed by 6 trials where they were decremented, and so forth. For each horizontal distance, 16 possible motion direction judgments were obtained. These were used to generate a psychometric function, which plotted the percentage of time the observer saw horizontal motion as a function of horizontal distance. Then, using probit analysis (Finney, 1971), the horizontal distance (HD_{50}), which led to a

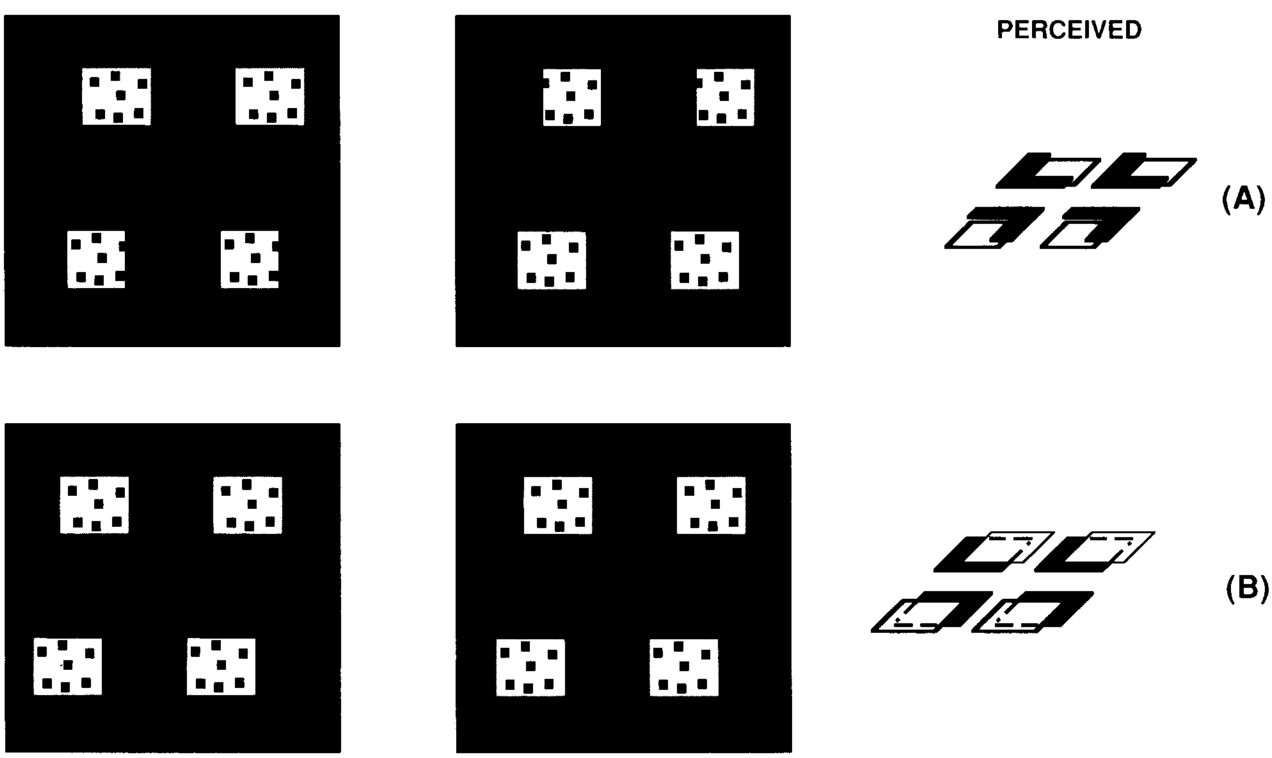

FIGURE 3. Illustration of the motion stimuli used in the first experiment. The two left and middle paneled stereograms (in box) are designed for convergent fusers. The perceived shapes (upon fusion) are illustrated in the right panel, where the black L-shaped tokens are seen in front (top box), and in back (bottom box). Note that in the back case (bottom box), the black L-shaped tokens appear extended and form larger square-like surfaces, which are "amodally" completed behind the white squares.

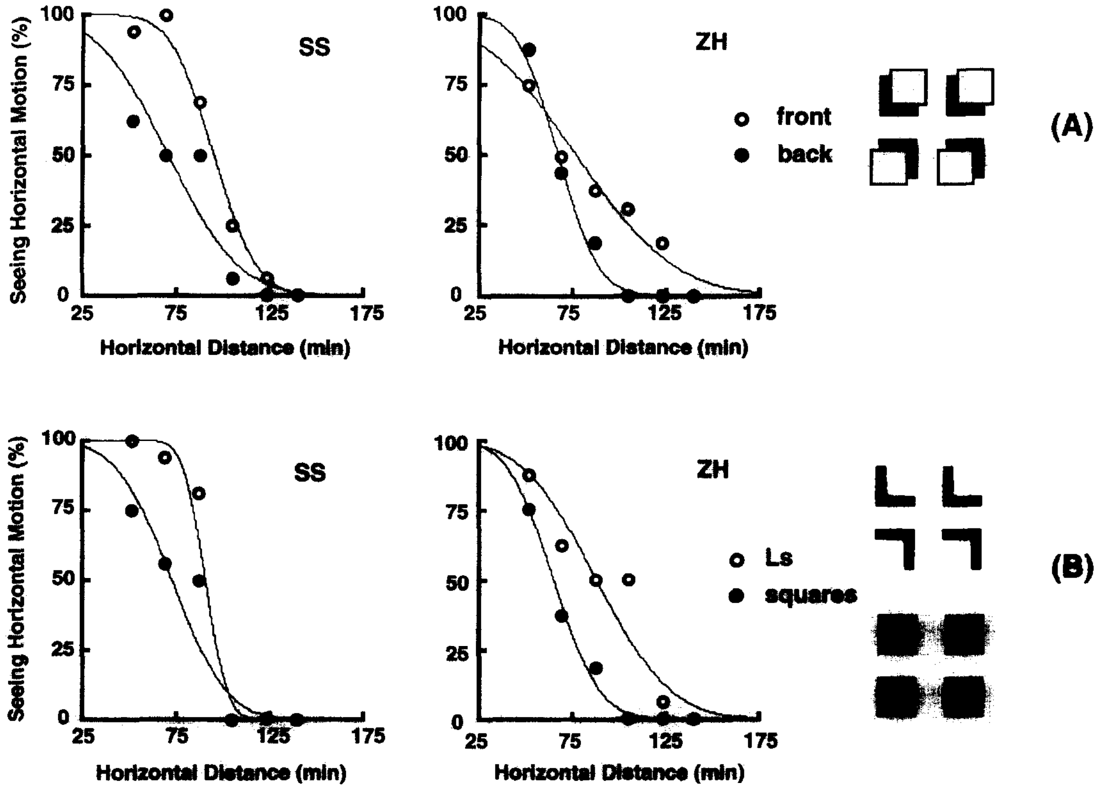

FIGURE 4. Comparison of results from Experiments 1 (A) and 2 (B), for observers SS and ZH. The drawings on the right side of the graphs depict the motion stimuli used in the experiments. (A) In Experiment 1, the HD_{50}'s for each observer were as follows: SS, HD_{50}(front) = 94.6 min arc, HD_{50}(back) = 70.2 min arc; for observer ZH, HD_{50}(front) = 77.0 min arc, HD_{50}(back) = 67.3 min arc; for observer KN (data not shown), HD_{50}(front) = 103.4 min arc, HD_{50}(back) = 93.3 min arc. (B) Meanwhile in Experiment 2, the HD_{50}'s for the same observers were: SS, HD_{50}(L) = 90.2 min arc, HD_{50}(S) = 72.7 min arc; for observer ZH, HD_{50}(L) = 87.4 min arc, HD_{50}(S) = 64.3 min arc; for observer KN (data not shown), HD_{50}(L) = 104.4 min arc, HD_{50}(S) = 89.4 min arc. Notice the consistency between the HD_{50}'s (front vs L, and back vs Square) for each subject over the two experiments. This indicates that motion correspondence was dependent on the shape similarity between the motion tokens as detailed in the text.

50% frequency of seeing horizontal motion was computed. This HD_{50} provides a preference index for perceiving horizontal motion in the display.

Six observers including the two authors (KN & ZH) and four naive observers (GP, RM, SS, & TW) with normal or corrected to normal vision participated in the first experiment. In addition, KN participated in experiment 2, and SS and ZH in Experiments 2, and 3. To reduce possible effects of observers' hysteresis during the experiments, all observers had about 100–150 practice trials before each test session (Ramachandran & Anstis, 1983; Shimojo & Nakayama, 1990).

RESULTS

Experiment 1. The Role of Amodal Surface Completion

The aim of this experiment was to pit the early filtering hypothesis against the surface representation hypothesis by using the motion stimuli illustrated in Fig. 3. As discussed earlier, the early filtering hypothesis predicts a similar horizontal motion preference in both the front and back cases, whereas the surface representation hypothesis predicts a greater horizontal motion bias in the front case.

Stimuli

The L-shaped motion tokens appeared in black (0.03 cd/m^2), and were attached to white squares (85 cd/m^2), displayed against a gray background (42 cd/m^2). When these tokens were arranged to be in front of their adjoining squares, the sizes of their horizontal and vertical limbs were 37.9 × 7.7 min arc and 11.2 × 38.7 min arc, respectively. When they were arranged in back, the horizontal widths of their vertical limbs changed as a function of their disparity magnitude in each binocular half-image. However, the average horizontal width between the left and right eyes always remained the same as the horizontal width in the front condition. The sizes of the adjacent white squares were 33.5 × 31.0 min arc when they appeared in front of the L-shaped tokens. When these white squares were seen behind the L-shaped tokens, their widths changed in accordance with the magnitude of the binocular disparity. Meanwhile, the binocular disparity between the L-shaped tokens and their adjacent squares was 8.9 min arc each. Finally, the center-to-center vertical distance between the L-shaped tokens was kept the same throughout the experiment (121.3 min arc).

Results

The data from observers SS and ZH are shown in Fig. 4(A). The filled and open circles represent the mean percentage for seeing horizontal motion at a given horizontal distance for the back and front cases, respectively. These data were further fitted with Z-score curves (probit analysis) as shown in the same figure. Note that the data obtained for the front condition shifts towards the right relative to the one for the back condition, indicating that there is a horizontal motion bias in the front condition. The average HD_{50}s of all six observers were 91.4 ± 5.0 min arc for the front case and 77.4 ± 7.4 min arc for the back case. The difference between these two conditions was significant ($t = 3.611$, $P < 0.01$). Thus, this result confirms the surface hypothesis which predicts that a lesser horizontal motion bias would be observed in the back condition, and argues against the early filtering hypothesis which predicts the same horizontal motion bias in both depth conditions.

In addition to the difference in perceived motion bias, all six subjects also reported another difference between the two conditions. In the front case, as the L-shaped tokens were seen translating in the vertical direction, they were simultaneously perceived to rotate in 3-D space. However in the back case, no rotation was perceived. Such a difference in the perceived motion paths, implies very different sets of motion correspondence rules used to match the edges of the tokens. In the front case, the horizontal and vertical limbs of the L and reversed L-shaped tokens corresponded to each other when a vertical motion was perceived [Fig. 3(A)]. However, in the back case, the visible part of an amodally completed square (i.e. the physical limbs of the L/reversed L-shaped tokens) corresponded to the invisible part of another amodally completed square (i.e. a physically non-existent part of the tokens) [Fig. 3(B)]. Hence, this phenomenology of the motion path pattern also provides an additional support for the view that the matching process occurs at the surface representation level.

Experiment 2. L/Square Control Experiment

Our support for the surface hypothesis hinges on the assumption that in the back case, the L-shaped tokens were perceived as amodally completed into square-shaped tokens. As such, these tokens would perceptually be rendered as the same shape (square). So unlike the front case, where the shapes of the L and reversed L tokens preserved an orientation (shape) difference, a horizontal motion bias would not be expected in the back case. In the current control experiment, we explicitly tested the assumption that our results in experiment 1 were due to shape dissimilarity in the front case (L vs reversed L), and shape similarity in the back case (square vs square). Instead of relying on an amodally completed square, we used a real square shaped token and simply compared it to a real L shaped token. In this situation, no adjoining squares were used [Fig. 4 (B)]. If our assumption of the roles of form similarity in 3-D is correct, then we would expect the findings in the current experiment to be similar to the one in Experiment 1. Perceived real squares just like perceived amodally completed squares would not show as much horizontal motion bias.

Stimuli

The motion square-shaped tokens that replaced the illusory squares were about 37.9×38.7 min arc, which was about the sizes of the amodally completed squares in the back case of Experiment 1.

Results

Figure 4(B) depicts the data of observers SS and ZH. Filled and open circles represent the percentage of times the observers saw the square and L-shaped tokens moved in the horizontal direction. The data for the L-shaped token condition (open circles) predictably shifts rightwards relative to the ones for the square-shaped token condition (solid circles). The shift is of approximately the same magnitude as seen in front case in Experiment 1, whereas the HD_{50} for the real squares was of approximately the same as seen with amodal completion in Experiment 1. This indicates a motion bias in the horizontal direction for the L-shaped token condition. The data from observer KN, not shown here, also demonstrated a similar tendency. Furthermore, as in Experiment 1 subjects observed the same 3-D rotational motion as the L-shaped tokens moved in the vertical direction. Hence, the findings of the current experiment supports the assumption that the form similarity among the real squares, and the form similarity of the amodal squares is more or less equivalent.

Experiment 3. The Oriented Bar Experiment

(A) Amodal surface completion

A concern that could be raised in the above experiments is that the L-shaped tokens used were possibly ineffective in isolating the early filtering mechanisms because such mechanisms do not have L-shaped receptive fields. This criticism is somewhat far fetched, however, because L and reversed L-shaped tokens can activate populations of early level oriented filters differentially. Nevertheless, we decided to provide further support for our conclusions, by using stimuli that more directly simulate the shapes of early cortical receptive fields, i.e. by using 45° oriented bar elements (Fig. 5). For the reasons stated earlier, the surface hypothesis predicts that when the oriented bars are seen in front of their adjacent and stationary green diamonds, there should be a horizontal motion bias as the top and bottom tokens differ in their orientations (45 vs $-45°$). However, this horizontal motion bias will diminish when all the bars are seen as amodally completed squares occluded by their adjacent diamonds in back.

Stimuli. The motion stimuli are illustrated in the top (front case) and bottom (back case) stereograms of Fig. 5 (not drawn to scale). The approximately 45° oriented white bar motion tokens ($81.2\ cd/m^2$) were attached to

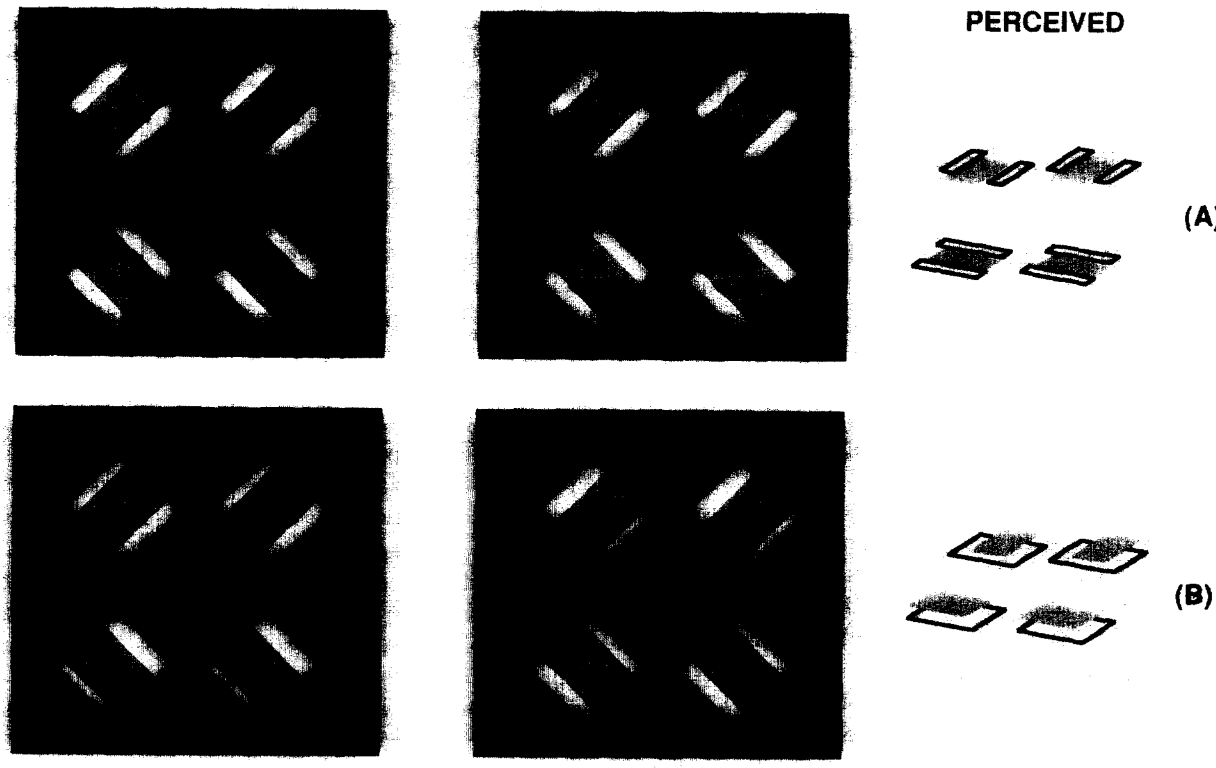

FIGURE 5. Illustration of the motion stimuli used in Experiment 3(A). The two left and middle paneled stereograms (in box) are designed for convergent fusers. The perceived shapes are illustrated in the right panel, where white 45° oriented bars are seen in the front case (top box), and in the back case (bottom box). Note that in the bottom box, the white bars appear extended and form larger diamond-like surfaces, which are "amodally" completed behind the green squares.

their adjacent green diamonds (67.4 cd/m^2, size of 43 × 41 min arc) and presented against a gray background (15.1 cd/m^2). The sizes of the white bars were 38.4 × 10.0 min arc (edges) in the front case. When the white bars were seen behind the green diamonds in the back case, their sizes were different for each pair of bars depending on the binocular disparity; however their average sizes remained at 38.4 × 10 min arc (edges) as in the front case. The motion tokens' center-to-center vertical distance (108 min arc) was kept the same during the experiment. The disparity between the white bars and their adjacent diamonds was about 6.7 min arc each. As in the previous experiments, we used a frame presentation duration of 300 msec. In addition, we also measured motion perception at a 150 msec duration.

Results. Figure 6 shows the percentage of times observers (SS & ZH) saw horizontal motion at given horizontal distances for the front (open symbols) and

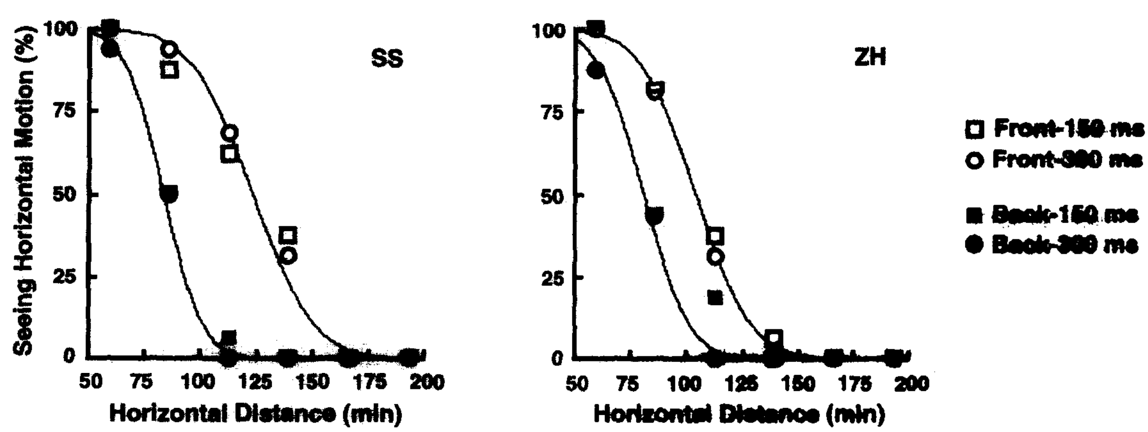

FIGURE 6. The proportion of seen horizontal motion as a function of horizontal distance in Experiment 3(A). The data obtained from the front and back cases are represented by the open and solid symbols, respectively. The data for the stimulation duration of 300 msec (circles), are also fitted by solid curves obtained from probit analysis. Clearly, the data from the back cases shift leftward, which indicates less horizontal motion bias was perceived in the back case. The computed HD_{50}'s for each observer were as follows: SS: 120.1 min arc (front–150 msec); 123.3 min arc (front–300 msec); 88.0 min arc (back–150 msec); 83.5 min arc (back—300 msec); ZH: 105.5 min arc (front–150 msec); 104.1 min arc (front–300 msec); 89.5 min arc (back–150 msec); 80.9 min arc (back–300 msec).

back cases (solid symbols), at the two frame presentation durations. Note that the different durations produced similar motion perception (150 msec: squares; 300 msec: circles). Furthermore, the data obtained from the front cases shift leftward relative to the data from the back cases, indicating a horizontal motion bias in the front cases. [The HD_{50} differences between the front and back cases were 39.8 min arc (300 msec) and 32.2 min arc (150 msec) for SS, and 23.2 min arc (300 msec) and 25.0 min arc (150 msec) for ZH.]

(B) Modal surface completion: the critical role of perceptual transparency

The surface hypothesis argues that the form similarity between the neural images of the motion tokens at the surface representation level plays a critical role in determining apparent motion correspondence. So far, this hypothesis has only been supported by instances where amodal surface completion occurred. To generalize our findings, we next considered an experiment which utilized another aspect of the surface completion, that of a modal surface completing in front.

Recent evidence indicates that modal and amodal surface completion share many similar characteristics (Kanizsa, 1979; Kellman & Shipley, 1991; Nakayama & Shimojo, 1992). As mentioned earlier, one example of modal surface completion is the perception of surface transparency which occurs only when the luminance conditions for its appearance is appropriate. In particular, Metelli (1974) has shown that in order for a region to be perceived as transparent, it must assume an intermediate luminance level relative to the surface region that it is presumed to occlude and the background. Like amodal surface completion, modal surface completion can also be enhanced or eliminated by manipulating binocular disparity (Nakayama & Shimojo, 1990; Nakayama *et al.*, 1990). Therefore, in the following experiments, we jointly manipulated relative luminance and binocular disparity to control perceived transparency. Because the surface completion associated with a transparent surface requires both disparity and luminance to be appropriate, we predicted that apparent motion correspondence would be altered only under this same set of restricted conditions.

The four stereograms in Fig. 7 illustrate the four motion stimuli used in the experiment. When fused convergently, the stereogram in condition (B) has its luminance and stereo depth values set to allow the reader to perceive red transparent square-shaped motion tokens that are located in front of black squares, i.e. modal completion in the front plane. But in the remaining conditions (A), (C), and (D), the luminance and stereo depth values are invalid for seeing transparency, i.e. modal surface completion does not occur. Under these three latter conditions only opaque oriented bars are seen, not transparent squares. Consequently, only in the transparent case (B) has modal surface completion effectively reduced the surface shape difference between top and bottom tokens. Correspondingly, oour prediction is that only in the transparent case where modal completion has occurred will the horizontal motion bias be reduced.

Stimuli. The motion stimuli were similar to those illustrated in the stereograms of Fig. 7 (not drawn in scale). The motion tokens were 45° oriented bars, presented against a white background (81.4 cd/m^2). The sizes of the motion tokens in conditions (A) and (C), and conditions (B) and (C), were the same as the front and back cases in the previous amodal surface completion experiment (Fig 5). The luminance of the black parts was 0.01 cd/m^2, and of the red parts was 161 cd/m^2 (0.619, 0.347). The disparity between the oriented bars and their adjacent diamonds was 67 min arc. The motion tokens' center-to-center vertical distance (108 min arc) was kept the same during the experiment.

Results. Figure 8 shows the percentage of times observers (SS & ZH) saw horizontal motion at the given horizontal distances for all conditions. The solid circles representing the transparent case [Fig. 8(B)], shift leftward relative to the other cases [non-transparent conditions, Fig. 8(A), (C), and (D)]. This indicates less horizontal motion bias in the transparent case. Such a finding is similar to the previous results utilizing amodal surface completion phenomenon in back. Both show that when shape differences at the surface representation level are reduced, apparent motion correspondence strength is predictably altered. The similar results from both modal and amodal completion experiments further rule out the possibility that depth alone is responsible for the change in motion correspondence because each occurs at an opposite sign of depth. Thus, together, they provide strong support for the surface hypothesis: that perceived surface shape determines the strength of apparent motion correspondence.

DISCUSSION

Several studies have also reported that apparent motion correspondence strength can be influenced by the surface properties of motion stimuli, such as convexity/concavity, illusory shape, occlusion etc. (Ramachandran, 1988; Ramachandran, Inada & Kiama, 1986; Shimojo & Nakayama, 1990). These findings are important as they imply that in addition to the early filtering level, the outputs from a later surface representation stage also contribute to the motion correspondence process. Our current experiments provide further insight into apparent motion perception by demonstrating that the motion bias for stimuli with similar early filtering properties can be altered by changing the motion tokens' surface shape properties. Thus it is likely that apparent motion correspondence is mainly determined by the neural images at the surface representation level, rather than the ones at the early filtering level.

The surface representation level compared to the early filtering level, operates at a relatively large spatial scale and is more concerned with the surface formation of objects, rather than local features of objects. Thus, from a perspective of coding efficiency, the neural images at

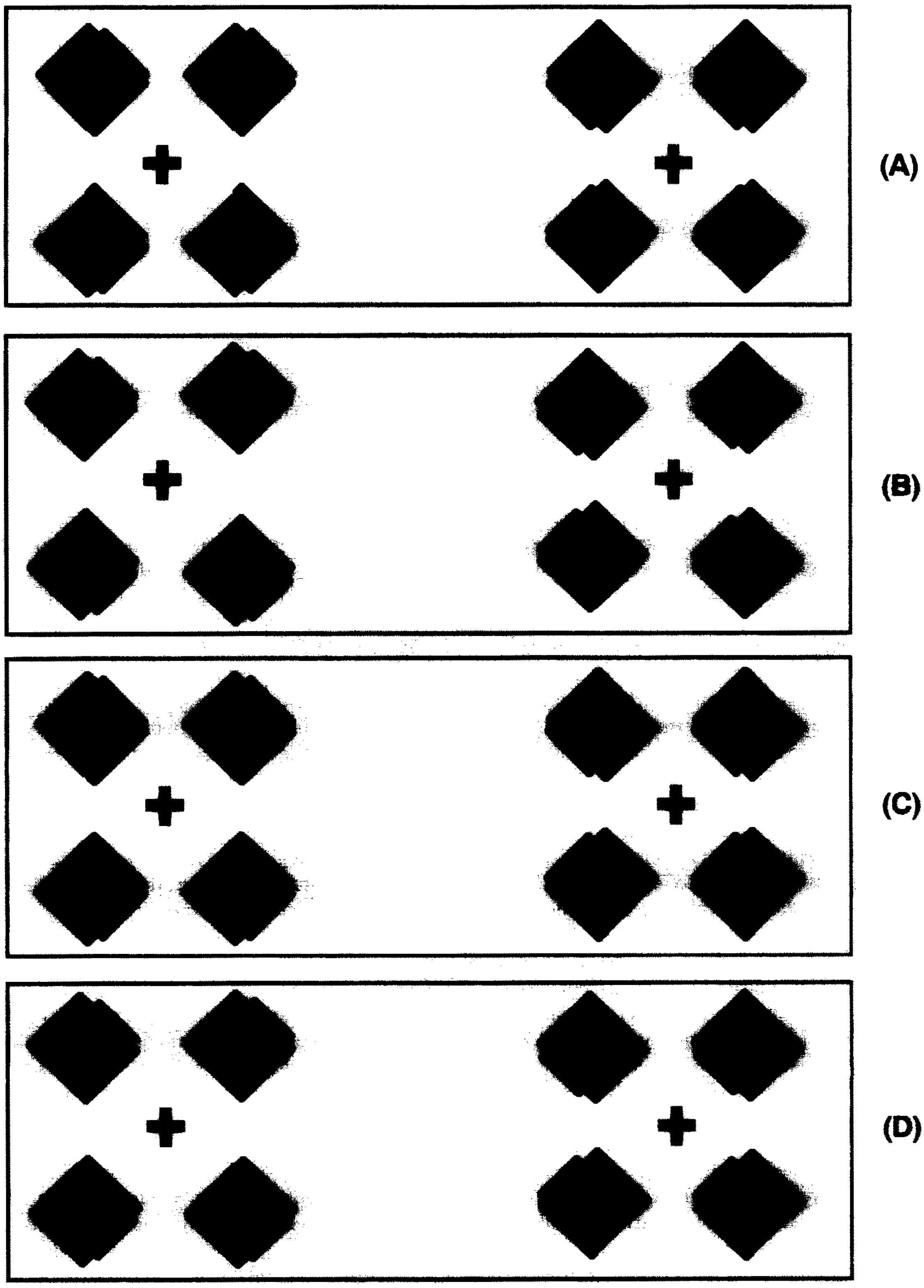

FIGURE 7. Illustration of the motion stimuli used in the modal surface completion Experiment 3(B). The four paneled stereograms (in box) are designed for convergent fusers. The transparent case is seen in condition (B) (the second panel). No transparency is perceived in the remaining three stereograms for various reasons: (A) invalid stereo arrangement; (D) invalid luminance condition; (C) invalid stereo arrangement and luminance condition.

the surface representation level may have an advantage in labeling an object in motion as a whole (such as a partially occluded moving object), over the neural images at the early filtering level. For example, motion energy mechanisms (Adelson & Bergen, 1985) are likely to yield very spurious signals for targets which become occluded during a motion trajectory. As such, these mechanisms would not be appropriate to sense the motion of an object or surface. This integrative characteristic of the surface representation level is consistent with the impression that the perception of apparent movement is about the motion of the whole object, rather than its attributes.

Again, the important role of the surface representation level in apparent motion arises from the fact that motion occurs in a 3-D world and that an object can become occluded during part or all of its trajectory. In the natural scene, a moving object could disappear and reappear when it passes behind an opaque object, leaving an invisible motion track as it does. As such, the apparent motion perception could be a consequence of the visual system attempting to interpolate within this invisible track. Thus, when tracking a moving object with a partially occluded part in 3-D space, labeling the object's surface properties including the occluded part by the surface completion mechanisms is more reliable than labeling it by the simple local feature mechanisms.

Still unresolved is whether a yet higher level of explicit object representation also participates in solving the apparent motion correspondence problem. We speculate that apparent motion correspondence is probably largely mediated at the level of surface representation for two reasons. First, because objects and their relation to each other in our visual world are normally defined by surface discontinuations (Gibson, 1979), a description at the surface representation level is sufficient to label the moving objects. Second, because higher order representations may take more processing capacity (attention), they may not be suitable for monitoring (fast) moving objects.

Minimum motion hypothesis

Recall that in Experiment 2 when the L-shaped tokens in the front case were seen translating in the vertical direction, a rotating motion of the individual token was simultaneously experienced. Such a translating plus rotating motion however, was not seen with square-shaped motion stimuli (see also Shimojo & Nakayama, 1990). This additional motion pattern in our experiment may explain a horizontal motion bias (or less vertical motion bias) in the front case, in accordance with the minimum motion hypothesis (Foster, 1978). This hypothesis states that if multiple elements are presented at different times, the visual system has a tendency to perceive apparent motion between the motion tokens which make the minimum shape transformation and travel the shortest motion path. If this hypothesis is correct, then our findings may suggest that the process of determining the least motion operates at the level of surface representation, or higher. Conversely, the minimum motion hypothesis argues that two motion tokens with similar surface shapes will have a stronger motion correspondence because the resulting motion requires less of a shape transformation. In other words, the motion correspondence strength between two motion tokens may reflect the internal effort of the brain to transform the surface shapes of tokens in motion.

The perceptual and phenomenological primacy of surfaces

The conclusion drawn from the current apparent motion experiments echoes those from our previous experiments on visual search and visual texture discrimination (He & Nakayama, 1992, 1993b). In those experiments, we also manipulated the binocular disparity of stimulus elements and found that visual search and texture discrimination performances were impaired when surface completion mechanisms resulted in the two elements' surface shapes becoming less distinct. Furthermore, we also showed that when the surface completion mechanisms caused the two elements' surface shapes to be more distinct, visual search performance became faster (He & Nakayama, unpublished results). Based on these and other observations, we concluded that in rapid texture discrimination and visual search tasks, the visual system cannot ignore the information related to surface layout (He & Nakayama, 1992, 1993b).

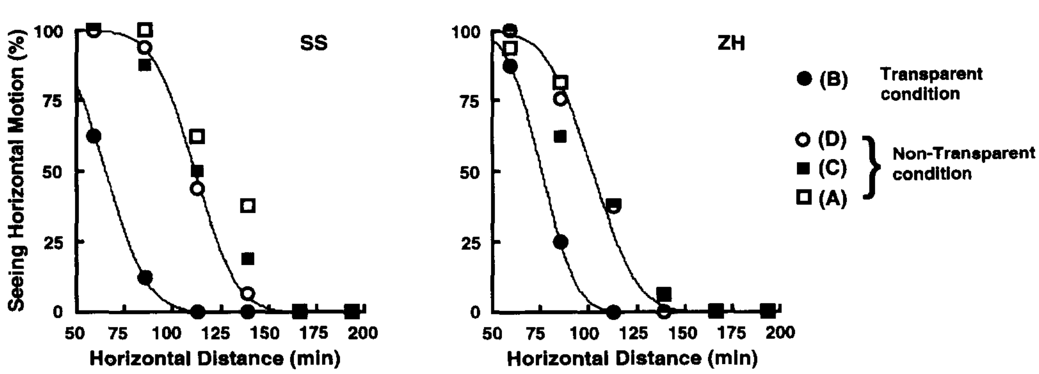

FIGURE 8. The proportion of seen horizontal motion as a function of horizontal distance, for each of the four conditions in Experiment 3(B). The solid and open circles represent the data for conditions (B) and (D), while the solid and open squares for conditions (C) and (A), respectively. The data for conditions (B) and (D) are fitted by solid curves obtained from probit analysis. Note that the curve for the transparent condition (B) shifts leftward indicating less horizontal motion bias. The computed HD_{50}'s for each observer were as follows: SS: 65.1 min arc (B); 126.9 min arc (A); 114.5 min arc (C); 111.6 min arc (D); ZH: 75.6 min arc (B); 104.9 min arc (A); 102.2 min arc (C); 101.9 min arc (D).

The commonalty seen between such a wide variety of visual tasks is striking. In each case, surfaces rather than features determines the outcome of the experiment. It indicates that surfaces rather than features comprise the input or raw material upon which the mechanisms of visual search, texture segregation and apparent motion must operate.

What is also of interest is that those visual tasks where we have shown that surfaces play a decisive role are also the same tasks traditionally thought to be fairly closely related to low level vision. Visual search, visual texture segregation and apparent motion have all been considered as lower level visual operations, so low that previous researchers (Julesz, 1981; Treisman & Gelade, 1980; Ullman, 1979) have suggested 2-D low level inputs for such mechanisms. Although our results contradict these assumptions, we do not wish to abandon the idea that these surface mechanisms are still nonetheless fairly primitive, say in relation to visual object recognition. Visual search (particularly easy search tasks), texture segregation, and apparent motion appear as more or less autonomous processes, not requiring scrutiny or high level knowledge. More important, what also characterizes these processes is their speed or immediacy. This leads us to suggest that visual surface representation, an inherently depth dependent process, comprises the most primitive visual representation upon which other very rapid visual mechanisms must depend.

Also of major interest is the fact that this is the same level of representation which characterizes our immediate perceptual phenomenology. When presented with a display of two bars separated by an image patch in front, our first impression is that of a single occluded square, not two separate bars in back (see Fig. 5). When presented with two red bars in front, separated by a black image patch in back (as in Fig. 7), we do not see two separate bars, but perceive these bars to be part of a larger transparent surface in front. What we take pains to note here is that what we "see" and what we are immediately conscious of in the display also determine the outcome of well controlled perceptual experiments.

These general conclusions are very different from those ordinarily derived from visual psychophysical experiments. In visual psychophysics, the outcome of well controlled experimentation, if successful, is usually and approvingly understood in terms of the properties and interactions of mechanisms which have no direct counterpart in conscious perception. For example, the detection of colored lights presented on a white background is dependent on a mechanism which subtracts medium from long wavelength cones; the detection of gratings is determined by the adaptation state of spatial frequency channels, etc. In our experiments, which incidentally we think have comparable methodological objectivity—we ask observers to do specific visual tasks, we do not ask for subtle phenomenological impressions—we are forced to resort to a very different set of explanatory entities, a level of surface representation which corresponds to our immediate consciousness, to our "seeing".

We note therefore, the possible existence of a significant conceptual break or dichotomy between those aspects of vision dependent or built upon unseen (or unconscious) vs and those aspects of vision dependent upon seen (or conscious) properties. Our identification of the surface representation level with conscious awareness is not new. It was originally outlined by Jackendoff (1987) and further elaborated by Crick and Koch (1992). Whether such a distinction will have wide impact or additional explanatory power is unclear. What is clear, however, is that it reinforces our view that a visual surface representation is distinct from early level processing.

REFERENCES

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

Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London, Series B, 290*, 153-168.

Braddick, O. J. (1980). Low level and high-level processes in apparent movement. *Philosophical Transactions of the Royal Society of London, Series B, 290*, 137-151.

Burt, P. & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review, 88*, 171-175.

Crick, F. & Koch, C. (1992). The problem of Consciousness. *Scientific American, September*, 152–159.

Finney, D. J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.

Foster, D. H. (1978). Visual apparent motion and the calculus of variations. In Leeuwenberg, E. L. & Buffart, H. (Eds), *Formal theories of visual perception* (pp. 67–82). New York: Wiley.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research, 26*, 599–607.

He, J. Z. & Nakayama, K. (1992). Surfaces vs. features in visual search. *Nature, London, 359*, 231–233.

He, J. Z. & Nakayama, K. (1994). Apparent motion determined by surface layout, not by disparity or 3-D distance. *Nature, London, 367*, 173-175.

He, J. Z. & Nakayama, K. (1994). Perceiving textures: Beyond filtering. *Vision Research, 34*, 151-162.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, Mass.: MIT Press.

Julesz, B. (1981). Textons, the elements of texture perception and their interactions. *Nature, London, 290*, 91–97.

Kanizsa, G. (1979). *Organization in vision: Essays in Gestalt perception*. New York: Prager.

Kellman, P. J. & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23*, 141-221.

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

Michotte, A. (1954). *La perception de la Causalite*. Louvain: Publications Universitaires.

Nakayama, K. & Shimojo, S. (1990). Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposium on Quantitative Biology, 40*, 911-924.

Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surface. *Science, New York, 257*, 1357-1363.

Nakayama, K., Shimojo, S. & Ramachandran, V. S. (1990). Transparency: Relation to depth, subjective contours, luminance, and neon color spreading. *Perception, 19*, 497-513.

Prazdny, K. (1986). What variable control (long-range) apparent motion? *Perception*, *15*, 37–40.

Ramachandran, V. S. (1985). Apparent motion of subjective surfaces. *Perception*, *14*, 127–134.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, London*, *331*, 163–166.

Ramachandran, V. S. & Anstis, S. M. (1983). Perceptual organization in moving patterns. *Nature, London*, *304*, 529–531.

Ramachandran, V. S., Inada, V. & Kiama, G. (1986). Perception of illusory occlusion in apparent motion. *Vision Research*, *26*, 1741–1749.

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

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass: MIT Press.

Acknowledgement—This research was supported by the AFOSR Grant F4962-92-J-0016.