Vision Res. Vol. 31, No. 5, pp. 859–876, 1991
Printed in Great Britain. All rights reserved

0042-6989/91 $3.00 + 0.00

THE KINETIC DEPTH EFFECT AND OPTIC FLOW—II. FIRST- AND SECOND-ORDER MOTION

MICHAEL S. LANDY,[1] BARBARA A. DOSHER,[2] GEORGE SPERLING[1] and MARK E. PERKINS[1]
[1]Psychology Department, New York University, NY 10003 and [2]Psychology Department, Columbia University, NY 10027, U.S.A.

(*Received* 24 *August* 1989; *in revised form* 1 *May* 1990)

Abstract—We use a difficult shape identification task to analyze how humans extract 3D surface structure from dynamic 2D stimuli—the kinetic depth effect (KDE). Stimuli composed of luminous tokens moving on a less luminous background yield accurate 3D shape identification regardless of the particular token used (either dots, lines, or disks). These displays stimulate both the 1st-order (Fourier-energy) motion detectors and 2nd-order (nonFourier) motion detectors. To determine which system supports KDE, we employ stimulus manipulations that weaken or distort 1st-order motion energy (e.g. frame-to-frame alternation of the contrast polarity of tokens) and manipulations that create *microbalanced* stimuli which have no useful 1st-order motion energy. All manipulations that impair 1st-order motion energy correspondingly impair 3D shape identification. In certain cases, 2nd-order motion could support limited KDE, but it was not robust and was of low spatial resolution. We conclude that 1st-order motion detectors are the primary input to the kinetic depth system. To determine minimal conditions for KDE, we use a two frame display. Under optimal conditions, KDE supports shape identification performance at 63–94% of full-rotation displays (where baseline is 5%). Increasing the amount of 3D rotation portrayed or introducing a blank inter-stimulus interval impairs performance. Together, our results confirm that the human KDE computation of surface shape uses a global optic flow computed primarily by 1st-order motion detectors with minor 2nd-order inputs. Accurate 3D shape identification requires only two views and therefore does not require knowledge of acceleration.

KDE   Kinetic depth effect   Structure from motion   Shape   Optic flow

INTRODUCTION

When a collection of randomly positioned dots moves on a CRT screen with motion paths that are projections of rigid 3D motion, a human viewer perceives a striking impression of three-dimensionality and depth. This phenomenon of depth computed from relative motion cues is known as the kinetic depth effect (KDE; Wallach & O'Connell, 1953).

What are the important cues that lead to a 3D percept from such a display? Is it motion, or are there other important cues? If it is motion, then what kind of motion detection system(s) are used to support the structure-from-motion computation? Is a computation of velocity sufficient, or are more elaborate measurements necessary, such as of acceleration? These are the questions that we address in this paper.

In a series of recent papers (Dosher, Landy & Sperling, 1989a, b; Sperling, Landy, Dosher & Perkins, 1989; Sperling, Dosher & Landy, 1990), we examined the cues necessary for subjects to perceive an accurate representation of a 3D surface portrayed using random dot displays. In each trial of a new shape identification task we devised, subjects view a random dot representation of one of a set of 53 3D shapes and identify the shape and rotation direction. Shape identity feedback optimizes the subject's ability to compute shape from each type of motion stimulus. For accurate performance, the task requires either a 3D percept or a subject strategy that uses 2D velocity information in a manner that is computationally equivalent to that required to solve for 3D shape (Sperling et al., 1989, 1990; see the discussion of expt 2, below).

We have shown that the only cue used for the perception of three-dimensionality in these displays is motion (Sperling et al., 1989, 1990). Further experiments determined that global optic flow is used rather than the position information for individual dots, since accuracy remains high when dot lifetimes are reduced to as little as two frames (Dosher et al., 1989b). In that paper, we concluded that the input to the KDE computation is an optic flow generated by a 1st-order motion detection mechanism, such as the Reichardt detector (Reichardt, 1957). Two manipulations that perturb 1st-order motion energy mechanisms—flicker and polarity alternation—also interfered with KDE (Dosher et al., 1989b). In polarity alternation, dots change over time from black to white to black on a gray background. When compared to dots that remain white, polarity alternation was equally or slightly more detectable in a detection task, was poorer but still well above chance in a discrimination of direction of motion task (computed, presumably, using tracking of the dots or using more elaborate, 2nd-order motion detection mechanisms) but was useless for tasks requiring KDE or motion segregation. These latter two tasks require the evaluation of velocity in a number of locations simultaneously (Sperling et al., 1989). Shape identification performance in a range of conditions was shown to be monotonic with a computed index of 1st-order net directional power in the stimuli (Dosher et al., 1989b). Hence, for sparse dot stimuli, KDE depends upon a simple spatio-temporal (1st-order) Fourier analysis of multiple local areas of the stimulus.

In this paper, we further examine and generalize the contributions of several types of motion detectors to the optic flow computations used by the structure-from-motion mechanism.

MOTION ANALYSIS MODELS AND THE KDE

1st-order motion analysis

To motivate the stimulus conditions studied here, we begin by summarizing models of early motion detection and analysis. Several recent motion detection models (van Santen & Sperling, 1984, 1985; Adelson & Bergen, 1985; Watson & Ahumada, 1985) share as a common antecedent the model proposed by Reichardt (1957). We refer to this class of models as 1st-order motion detectors. Below, 2nd-order mechanisms involving additional processing stages will be discussed. In the Reichardt detector, luminance is measured at two spatial locations *A* and *B*. The measurement at position *A* is delayed in time, and then cross-correlated over time with the measurement at position *B*, resulting in a "half-detector" sensitive to motion from position *A* to *B*. A second such "half-detector" sensitive to motion from *B* to *A* is set in opponency with the first, resulting in the full motion detector. van Santen and Sperling (1984, 1985) have investigated this model along with extensions involving voting rules for combining outputs of many detectors to enable predictions of psychophysical experiments, resulting in their Elaborated Reichardt Detector (ERD).

An alternative way of characterizing motion detection is in the frequency domain. A motion detector can be built of several linear spatio-temporal filters. Each filter is sensitive only to energy in two of the four quadrants in spatio-temporal Fourier space (ω_x, ω_t). In other words, the filters are not *separable*. Their receptive fields are oriented in space-time, and thus they are sensitive to motion in a particular direction and at a particular scale (Adelson & Bergen, 1985; Burr, Ross & Morrone, 1986; Watson & Ahumada, 1985). The Fourier "energy" (the squared output of a quadrature pair of filters) in each of two opposing motion directions is computed, and put in opponency. This "motion energy detector", proposed by Adelson and Bergen (1985), and the ERD differ in their construction and in the signals available at the subunit level, but are indistinguishable at their outputs (Adelson & Bergen, 1985; van Santen & Sperling, 1985).

The structure-from-motion computation relies upon the measurement of image velocities at several image locations. The KDE shape identification task that we use here can be solved by categorizing velocity at six spatial locations into three categories: leftward, approximately zero, and rightward (Sperling et al., 1989). Thus, in order to discriminate the 53 test shapes by KDE, motion detection must be followed by at least some rudimentary local velocity calculation.

In order to signal velocity, the outputs of more than one such 1st-order motion detector must be pooled. Speed may be computed by pooling only two detectors (a motion and a "static" detector, Adelson & Bergen, 1985). To signal motion direction, signals must be pooled across a variety of orientations (Watson & Ahumada, 1985). Finally, in order to solve the "aperture problem" for more complex stimuli (Burt & Sperling, 1981; Marr & Ullman, 1981), signals may be pooled over a variety of directions and perhaps scales (Heeger, 1987).

In the previous paper (Dosher et al., 1989b), shape identification performance was shown to relate directly to the quality of the signal available from 1st-order motion detection mechanisms. Each stimulus consisted of a large number of dots on a gray background representing a 2D projection of dots on the surface of a smooth 3D shape under rotary oscillation. In one condition (contrast polarity alternation), the dots were first brighter than the background ("white-on-gray"), then darker than the background ("black-on-gray"), then bright again, in successive frames. For a dense random dot field (50% black/50% white) under simple planar motion, polarity alternation causes a percept of motion opposite to the true direction of motion (the "reverse-phi phenomenon", Anstis & Rogers, 1975); reverse-phi is thought to reflect a spatio-temporal Fourier analysis of the stimulus, since contrast reversal reverses the direction of motion of the lowest-frequency Fourier components (van Santen & Sperling, 1984). With contrast reversal, the outputs of 1st-order motion detection mechanisms no longer simply signal the intended direction and velocity of motion. Contrast reversal stimuli do not yield a depth-from-motion percept (Dosher et al., 1989b). We take this as evidence that the KDE relies upon input from a 1st-order motion analysis.

2nd-order motion analysis

For the sparse random dot stimuli (Dosher et al., 1989b), contrast polarity alternation eliminated the perception of structure from motion. Nonetheless, subjects could judge the direction of patches of contrast polarity alternating dots undergoing simple translation. What kind of a motion detector might be used to correctly judge the motion of a translating, polarity-alternating dot? One simple possibility would be to first apply a luminance nonlinearity to the input stimulus. For example, if the input stimulus were full-wave rectified about the mean luminance, the polarity-alternating stimulus would be converted to the equivalent of rigid motion of a white dot on a gray background. Thus, a full-wave rectifier of contrast followed by a 1st-order analyzer (such as those discussed above) would be capable of analyzing such a motion stimulus correctly (Chubb & Sperling, 1988b, 1989a, b).

A motion detection system consisting of a contrast nonlinearity followed by a 1st-order detector is one example of a wide class of "2nd-order detection mechanisms", each of which consists of a linear filtering of the input (spatial and/or temporal), followed by a contrast nonlinearity, followed by a standard 1st-order motion detection mechanism. A number of results demonstrate the existence of both 1st- and 2nd-order motion mechanisms and show the contribution of both to the perception of planar motion (Anstis & Rogers, 1975; Chubb & Sperling, 1988b, 1989a, b; Lelkens & Koenderink, 1984; Ramachandran, Rao & Vidyasagar, 1973; Sperling, 1976).

Can both 1st- and 2nd-order motion mechanisms be used by the KDE system? The polarity-alternating dots did not yield an effective KDE percept of our 3D shapes. If one accepts the existence of both 1st- and 2nd-order motion mechanisms, why didn't the 2nd-order system support KDE? The KDE stimuli were relatively small (3.7 × 4.2 deg) and viewed foveally (eye movements were permitted throughout the 2 sec stimulus duration). Evidence from studies of planar motion suggests that both systems were available under these conditions (Chubb & Sperling, 1988b). For polarity alternation stimuli, the most salient low frequency components from the 1st-order system were in the wrong direction. We assume that the 2nd-order system yields a correct (if attenuated) analysis. Bad shape identification performance may have resulted either from the perturbed 1st-order analysis or because of competition between the 1st- and 2nd-order systems (which signaled opposite directions of motion in some frequency bands). Our evidence (Dosher et al., 1989b) demonstrated that 1st-order system input is the predominant input to KDE, but it did not exclude the possibility of input from 2nd-order motion detection mechanisms. To approach that question, we consider a KDE stimulus that produces a simple 2nd-order motion analysis, but to which the 1st-order motion system is, statistically, blind.

Microbalanced motion stimuli

Chubb and Sperling (1988b) defined a class of stimuli, called *microbalanced*, among which are stimuli with the properties that we desire. In expt 1 we concentrate on two examples of microbalanced motion stimuli. These stimuli are random in the sense that any given stimulus is a realization of a random process. As proven by Chubb and Sperling (1988b), if a stimulus is microbalanced then the expected output of every 1st-order detector (ERD or motion energy detector) will be zero. Thus, Chubb and Sperling defined a class of stimuli for which a consistent motion signal requires a 2nd-order motion analysis, and showed that the 2nd-order analysis predicted observers' percepts for several examples of the class.

The polarity alternation stimulus is not microbalanced; any given frequency band does show consistent motion, with the lowest spatial frequencies signalling motion in the wrong direction. This stimulus can be transformed into a microbalanced one as follows: for each dot, choose the contrast polarity randomly and independently for every frame. Any given 1st-order detector will be just as likely to signal rightward motion as it is to signal leftward motion since it will either see the same contrast polarity across any successive pair of frames or it will see contrast polarity alternate, with equal probability. One question we examine in this paper is whether the motion signal available from 2nd-order mechanisms can be used to compute 3D structure.

We present two experiments. In the first, we examine performance on a shape identification task for a variety of KDE stimuli. Several types of stimuli provide good 1st-order motion. Others are microbalanced and hence can only be analyzed by 2nd-order mechanisms. Still others offer good 1st-order motion, but involve camouflage similar to that available in some of the microbalanced conditions. We find that 1st-order motion is used, and that input from 2nd-order mechanisms may also be used but is not as robust. In a second experiment, we examine the residual shape percept from two-frame KDE stimuli in order to determine whether a single velocity field is a sufficient cue for shape identification or whether acceleration also is needed.

EXPERIMENT 1. POLARITY ALTERNATION, MICROBALANCE, AND CAMOUFLAGE

In the first experiment, a shape discrimination task is used with a variety of displays. First, in order to sensibly compare results to our previous work (Sperling et al., 1989; Dosher et al., 1989b), there are control conditions that are identical to those of our previous experiments (the "Motion without density cue, standard speed, standard intensity" and "Motion with polarity alternation, standard speed, standard intensity" conditions of the preceding paper). In addition to dots, randomly positioned disks and lines are also used here in order to examine the effects of the foreground token used to carry the motion. The disk and line tokens are larger than the single pixel dots, and hence have more contrast energy. They enable us to test whether our previous failure to find KDE with polarity alternation resulted from the low contrast energy in the stimulus. Two forms of microbalanced stimuli are used, allowing us to test KDE shape identification performance with stimuli to which 1st-order motion detectors are blind. Finally, we examine stimuli in which moving textured tokens are camouflaged by a similarly textured background.

Method

Subjects. There were three subjects in this experiment. One was an author, and the other two were graduate students naive to the purposes of this experiment. All had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the three subjects. These will be pointed out below.

White-on-gray dot stimuli. First, we briefly describe the stimuli that consist of bright dots moving on a gray background representing a variety of 3D shapes. This description will be somewhat abbreviated, since the same stimuli have been used in previous studies and more complete descriptions are available (Sperling et al., 1989). The other stimuli used in the present study result from simple image processing transformations applied to the white-on-gray dot stimuli.

Stimuli were based upon a fixed vocabulary of simple shapes consisting of bumps and concavities on a flat ground. The 3D shapes varied in the number, position, and 2D extent of these bumps and concavities. The process of generating the stimuli is illustrated in Fig. 1.

The first step in creating a stimulus involves the specification of a 3D surface. For a square area with sides of length s, a circle with diameter $0.9\,s$ is centered, and three fixed points, labeled 1, 2 and 3, are specified. For a given shape, one of two such sets of points is used (the upward-pointing triangle or the downward-pointing triangle, labeled u and d, respectively). The shape is specified as having a depth of zero outside of the circle. For each of the three identified points, the depth may be either $+0.5\,s$, 0.0, or $-0.5\,s$, which are labeled as $+$, 0, and $-$, respectively. The depth values for the rest of the figure were interpolated by using a standard cubic spline to connect the three interior points with the zero depth surround. Thus, there are 54 ways to designate a shape: u vs d, and for each of three interior points, $+$ vs 0 vs $-$. We designate a shape by denoting the triangle used, followed by the depth designations of the three points in the order shown in Fig. 1A. For example, $u-+0$

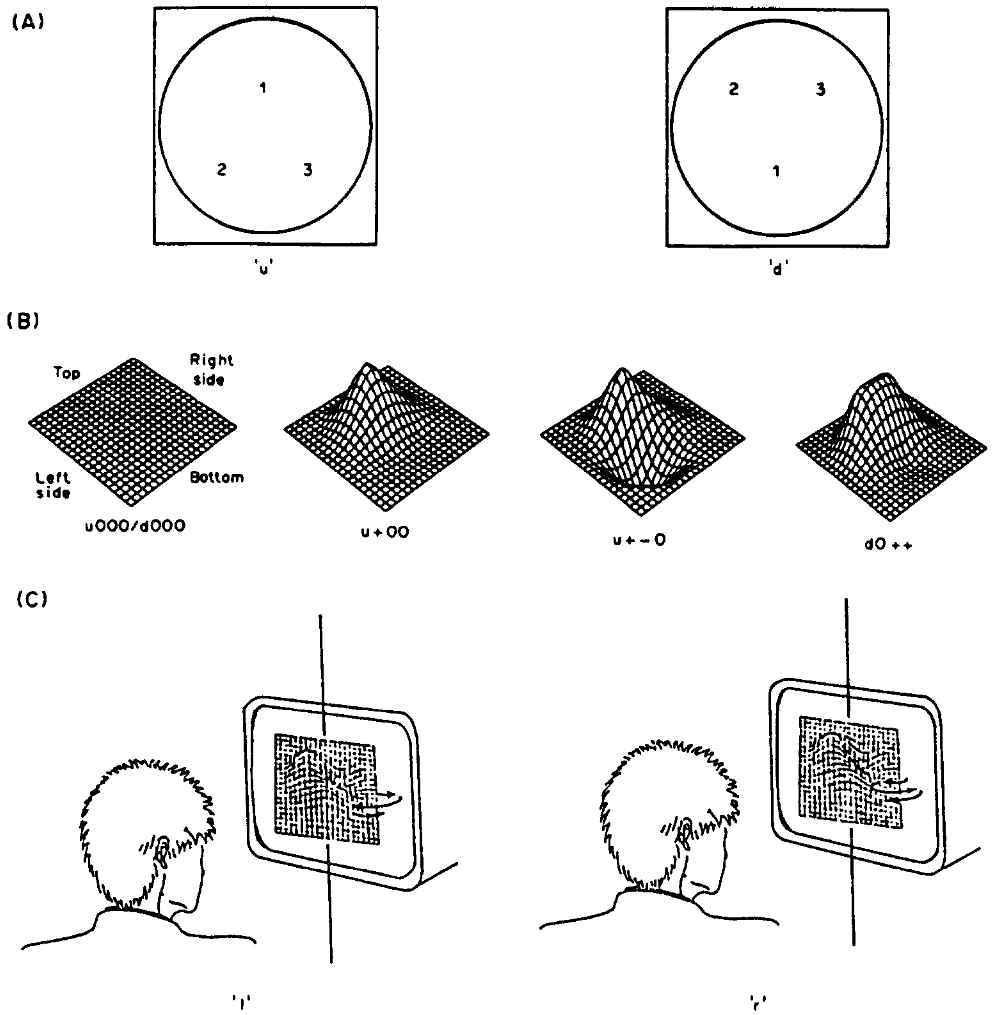

Fig. 1. Stimulus shapes, rotations, and their designations. (A) Shapes were constructed by choosing one of the two equilateral triangles represented here. Each point in the triangles was given a positive depth (i.e. toward the observer), zero depth, or negative depth, represented as +, 0 and −, respectively. A smooth shape splined these three points to zero depth values outside of the circle. A shape is designated by the choice of triangle (*u* or *d*), followed by the depth designations of the three points in the order given in the figure. (B) Some representative shapes generated by this procedure. All shapes consisted of a bump, concavity, or both, with a variation in position and extent of these areas. (C) Shapes were represented by a set of dots randomly painted on the surface of the shape, and wiggled about a vertical axis through the center of the display. The motion was a sinusoidal rotation that moved the object so as to face off to the observer's right, then his or her left, then back to face-forward (denoted *l*), or the reverse (denoted *r*).

is a shape with a bump in the upper-middle of the display, and a concavity in the lower-left (Fig. 1B). There are 53 distinct shapes, because *u*000 and *d*000 both denote a flat square.

Displays were generated by sprinkling dots randomly on the 3D surface generated by the spline, rotating that surface, and projecting the resulting dot positions onto the image plane using parallel perspective. A large number of dots are chosen uniformly over a 2D area somewhat larger than the s by s square, and each dot's depth is determined by the cubic spline interpolant (where the zero depth of the surround is continued outside the square). This collection of dots is rotated about a vertical axis that is at zero depth and centered in the display. The rotation angle $\theta(k)$ is a sinusoidal "wiggle": $\theta(k) = \pm 25\sin(2\pi k/30)$ deg, where k is the frame number within the 30 frame display. Thus, the display either rotated 25 deg to the right, then reversed its direction until it faced 25 deg to the left, then reversed its direction until it was again facing forward (labeled *l*), or rotated in the opposite manner (labeled *r*, see Fig. 1C). The displays presented these 3D collections of dots in parallel perspective as luminous dots (single pixels) on a darker background.

A stimulus name consists of the name of the shape followed by the type of rotation (e.g. $u + -0l$), resulting in 108 possible names. Using parallel perspective, there is a fundamental ambiguity with the KDE: reversing the depth values and rotation direction of a particular shape and rotation produces exactly the same display. In other words, a convexity rotating to the right produces exactly the same set of 2D dot motions as a concavity rotating to the left. Thus, $u + -0l$ and $u - +0r$ describe precisely the same display type. There is also no difference in display type among $u000l$, $u000r$, $d000l$ and $d000r$. This results in a total of 53 distinct display types.

These experiments used 54 white-on-gray dot displays, including two instantiations of the flat stimulus $u000$ (with different dot placements) and one instantiation of each other display type. Each set of dots was windowed to a display area of 182 × 182 pixels (corresponding to the $s \times s$ square), with dots presented as single luminous pixels.

When the dots on the surface of a shape move back and forth in the display, the local dot density changes as the steepness of the hills and valleys changes (with respect to the line of sight). In previous work (Sperling et al., 1989), we showed that this density cue is neither necessary nor sufficient for the perception of depth. However, it is a weak cue which one of three highly trained subjects was able to use for modest above-chance performance when it was presented in isolation. In other words, changing dot density is an artifactual cue to the task. As in previous experiments, we remove this cue by deleting or adding dots as needed throughout the display in order to keep local dot density constant. As a result of this manipulation, all displays had approx. 300 dots visible in the display window. The removal of the density cue results in a small amount of dot scintillation that neither lowers performance substantially nor appears to be useful as an artifactual cue (Sperling et al., 1989, 1990).

Other tokens. The 54 stimuli described so far consisted of luminous dots moving to and fro on a less luminous background. All other stimuli were based upon these displays. First, three conditions involved changes of the token that carried the motion. The moving dots were replaced with disks, patterned disks, or wires. We refer to the dot, wire, and disk conditions as *white-on-gray* stimuli, and the patterned disks as *pattern-on-gray*.

To create a disk stimulus, a dot stimulus is modified in the following way. Each luminous dot in the stimulus is replaced with a 6 × 6 pixel luminous diamond centered on the dot (Fig. 2b), which appears disk-like from the viewing distance used in the experiment. A sample image of white-on-gray disks is depicted in Fig. 2c, and is based on the white-on-gray dot stimulus frame shown in Fig. 2a.

The pattern-on-gray disk stimuli are generated in a similar fashion. The 6 × 6 diamond consists of 24 pixels which are a mixture of black and white (12 of each). These are displayed on an intermediate gray background. The diamond pattern and a sample stimulus frame are shown in Fig. 2d and e, respectively. Note that the diamond pattern has an equal number of black and white pixels in each row.

Other stimuli were based on "wires". Each dot was connected by a straight line (subject to the pixel sampling density) to all neighbors that were at a 2D distance no greater than 15.5 pixels (Fig. 2f). Note that a vector is drawn between two points based on their distance *in the image*, not on their simulated 3D distance. Since the lines were straight, when set in motion they objectively define a thickened surface with lines cutting through the interior of each bump and concavity. This may have yielded a perceived

Fig. 2 (*opposite*). Stimulus display generation for expt 1. (a) A single frame of a white-on-gray dots stimulus. All displays shown in this figure are based on this stimulus frame. (b) The diamond shape used to generate the disks from the dots. (c) A white-on-gray disks stimulus frame. (d) The patterned diamond for the pattern-on-gray condition. (e) A pattern-on-gray frame. (f) A white-on-gray wires frame. All pairs of dots in Fig. 2A were connected whose inter-point distance was less than 15.5 pixels. (g) A frame of dynamic-on-gray dots. In this condition each dot was painted black or white randomly and independently with probability of 0.5 for each color. (h) A frame of dynamic-on-gray disks. The same procedure as in (g) was applied to each pixel lying in each disk. (i) A frame of dynamic-on-gray wires. (j) A frame of dynamic-on-static disks. For both dynamic-on-static conditions (disks and wires), the tokens and the background consisted of random dot noise, and so the tokens cannot be discerned from a single static frame. (k) A frame of the pattern-on-static condition. This frame contains 300 copies of the pattern in (d) on a static noise background. The camouflage is quite effective. (l) An enlargement of the central portion of (k), with the patterned disks emphasized.

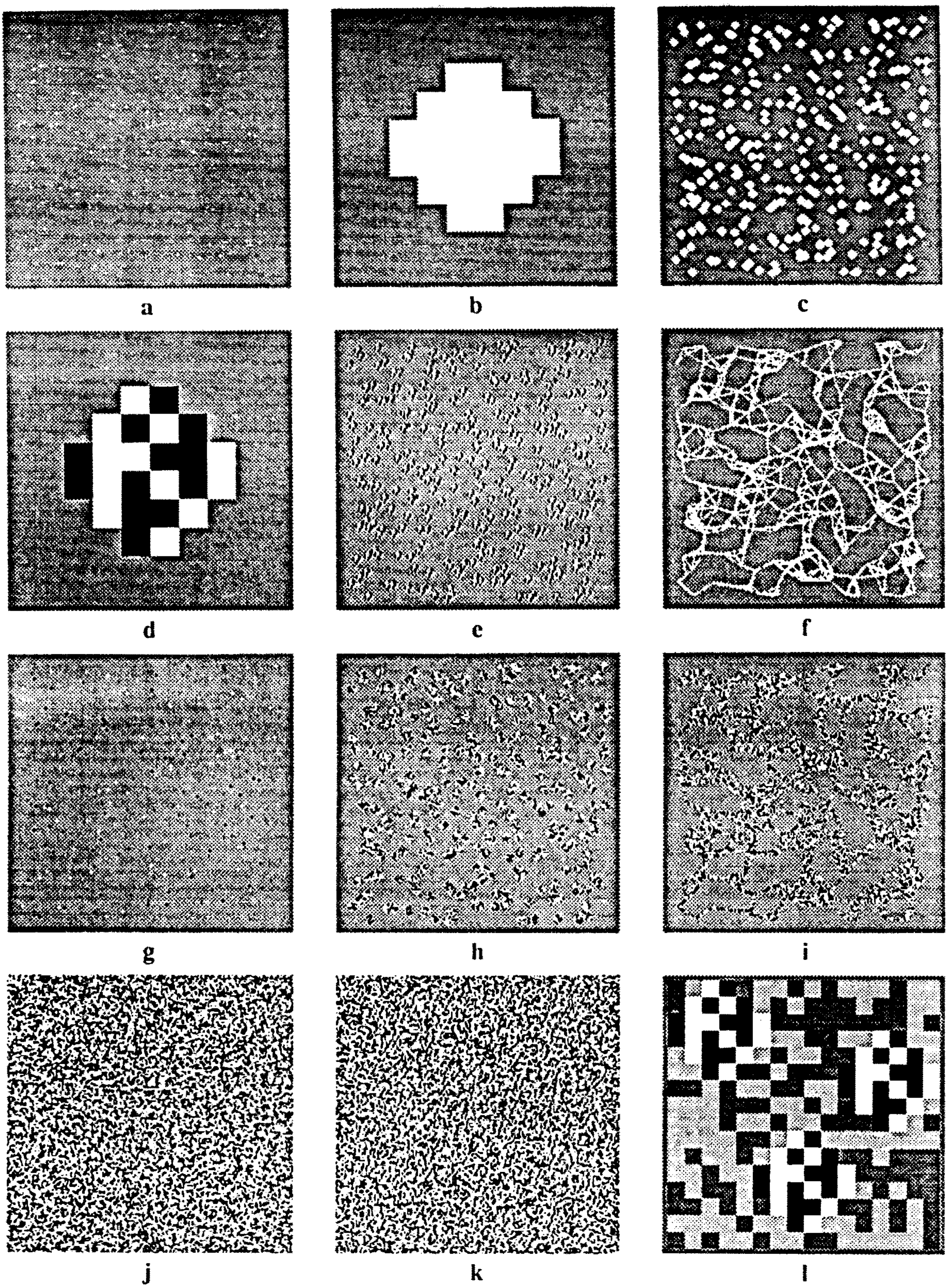

Fig. 2

(tesselated) surface having slightly less relative depth than the base surface. The choice of 15.5 pixels as the criterion for drawing a line was a compromise set in order to make sure that all stimulus dots became an endpoint to at least one line, and that no line was so long as to excessively cut through the simulated surface.

The white-on-gray disks and pattern-on-gray disks were based on the dot stimuli. The same exact instantiations were used in all three conditions. The *n*th frame of a given shape and rotation consisted of either dots, disks or patterned disks centered on the same set of image positions. For the wire stimuli, a new set of 54 instantiations was made.

Dynamic-on-gray. Three types of stimuli were used to explore the motion of patches of dynamic noise moving on a gray background. These stimuli are microbalanced, as we discussed in the previous section. These stimuli are derived from the dot, disk, and wire stimuli. To produce a dynamic-on-gray stimulus from a white-on-gray stimulus, simply change the luminance of each white pixel in each stimulus frame (i.e. the foreground or token pixels) to black randomly and independently with probability 0.5. Thus, foreground pixels undergo random contrast polarity alternation while background pixels are gray (i.e. have zero contrast). Sample frames are illustrated in Fig. 2g, h and i.

Dynamic-on-static. Two types of stimuli were used to explore the motion of patches of dynamic noise moving on a static noise background. This class of stimuli is also microbalanced (Chubb & Sperling, 1988b). We derive dynamic-on-static stimuli from the disk and wire stimuli. The foreground pixels consist of dynamic noise, just as in the previous dynamic-on-gray case. The background pixels consist of a static frame of patterned texture, where each pixel is randomly chosen to be either black or white with a probability of 0.5, just as the dynamic noise is. If a given pixel is a background position for two successive frames, then its color does not change. If that position is a foreground pixel in either or both frames, then there is a 50% chance that its color will change. A single frame of dynamic-on-static stimulus is simply a frame of random dot noise (Fig. 2j). The motion-carrying tokens are not discernible from a single frame. Rather, the areas of moving dynamic noise define the foreground tokens.

Contrast polarity alteration. Three stimulus conditions involved contrast polarity alternation. This stimulus manipulation was explored thoroughly for dot stimuli in the preceding paper (Dosher et al., 1989b). In this condition, the motion-carrying tokens alternate from white to black to white again on successive frames, all against a background of intermediate gray. Constrast polarity alternation was used with dots, disks, and wires, resulting in three polarity alternation conditions.

Pattern-on-static. The final condition involves pattern camouflage. This condition is derived from the pattern-on-gray stimuli. The gray background is replaced with a frame of static random dot noise. In other words, the patterned disk tokens move to and fro in front of a screen of static random dots, occluding it (and occasionally each other) as they pass by. A frame of this stimulus condition is pictured in Fig. 2k, and enlarged in Fig. 2l, where we have artificially highlighted the patterned disks for comparison to the pattern kernel shown in Fig. 2d. There are approx. 300 patterned disks in Fig. 2k. As you can see, the camouflage is quite effective. When the patterned disks move, as one might expect, they are easily visible (Julesz, 1971).

Display details. There are a total of 13 conditions (3 white-on-gray, 1 pattern-on-gray, 3 contrast polarity alternation, 3 dynamic-on-gray, 2 dynamic-on-static, and 1 pattern-on-static). There were 54 distinct displays for each of the 13 conditions. In all conditions, the displays are windowed to an area of 182 × 182 pixels. Displays were computed using the HIPS image processing software (Landy, Cohen & Sperling, 1984a, b), and displayed by an Adage RDS-3000 image display system.

Subjects MSL and JBL viewed these stimuli on a Conrac 7211C19 RGB color monitor. Only the green gun was used, and so stimuli appeared as bright green and black pixels (as dots, disks, lines or noise) on a green background of intermediate luminance. The stimuli subtended 3.7 × 4.2 deg. Stimuli were viewed monocularly through a dark viewing tunnel, using a circular aperture which was slightly larger than the stimuli.

Subject LJJ viewed the stimuli on a US Pixel PX15 black and white monitor with a P4-like phosphor. Here, stimuli subtended 2.9 × 2.9 deg, and appeared as white and black pixels on an intermediate gray background. Stimuli were viewed monocularly through a circular aperture in cardboard which approximately matched the hue of the displays, and which had approximately the same luminance as the stimulus background.

Each stimulus consisted of 30 stimulus frames. These were presented at a 60 Hz frame rate. Each frame was repeated four times, resulting in an effective rate of 15 new stimulus frames per second. Each stimulus lasted 2 sec. A trial sequence consisted of a fixation spot, a blank interval, the 30 frame stimulus, and a blank. The fixation and blank lasted either for 1 sec each (subjects MSL and JBL), or 0.5 sec each (subject LJJ). The background luminance remained constant throughout the trial sequence. Subjects were free to use eye movements to actively explore the display. Stimuli were viewed from a distance of 1.6 m. After each stimulus display, subjects responded with the name of the shape and rotation direction using either a computer keyboard or response buttons.

Slightly different image luminances were used for each subject. The background luminance for subjects MSL, JBL and LJJ were 31.0, 40.0 and 45.0 cd/m^2 respectively. Since isolated luminous pixels were used, the appropriate unit of measurement is *extra* μcd/pixel for bright pixels, and *removed* μcd/pixel for dark pixels, all at a specified viewing distance (Sperling, 1971). Stimuli were calibrated so that extra μcd/pixel and removed μcd/pixel were equal. For subjects MSL, JBL and LJJ, these were 13.2, 19.2 and 15.7 μcd/pixel, respectively, at a viewing distance of 1.6 m. Contrasts were nominally 100%.

Procedure. There were 13 stimulus conditions. For each condition, there were 54 stimuli (two instantiations of the flat stimulus *u*000, and one instantiation of each of the 52 other possible distinct shape/rotation combinations). This resulted in 702 stimuli, each of which was viewed once by each subject. These 702 trials were viewed in random order in six blocks of 117 trials. On a given trial, a stimulus was shown, subjects keyed in their responses, and then feedback was provided so that we measured the best performance of which the subject was capable. Each block lasted approx. 1 hr. Subjects ran several practice sessions on the white-on-gray dots condition before data were collected. Given the mix of stimuli in a given condition, guessing base rates for the identification of shape and rotation direction were between 1/53 (for a strategy of random guessing) and 2/54 (for a strategy of always answering *u*000*l*, or one of its equivalents).

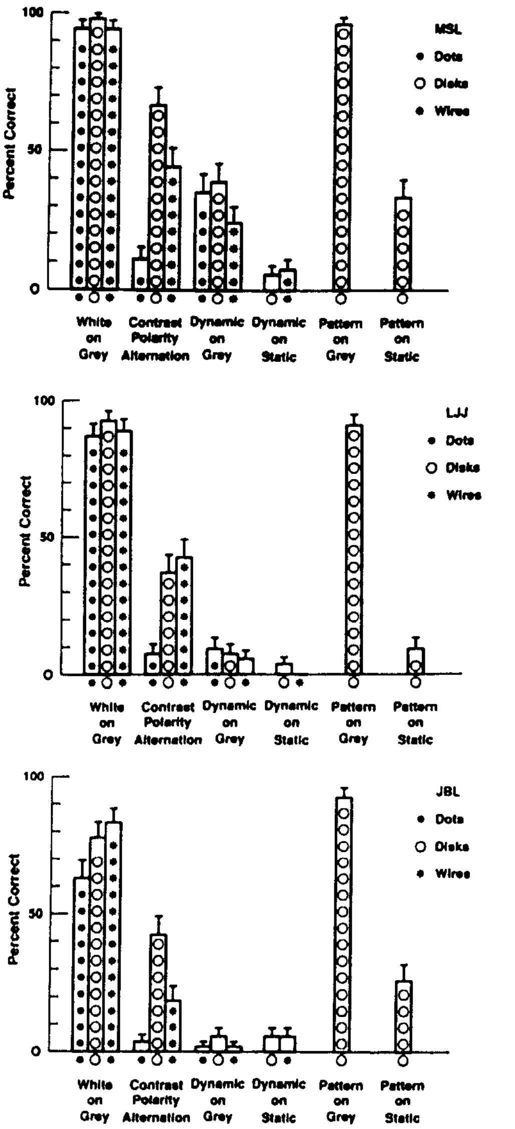

Fig. 3. Results of expt 1. Results are given for three subjects. Different symbols in the bars represent different tokens (large open dots for the disk and patterned disk tokens, small solid dots for the dot tokens, and asterisks for the wire tokens).

Results

The results for the three subjects are summarized in Fig. 3. Each performance measure given here is the percent correct over 54 trials. We discuss each class of stimulus condition in turn.

White-on-gray / Pattern-on-gray. As expected, the performance on the three white-on-gray and the one pattern-on-gray condition was uniformly high. The tokens provided excellent motion signals because they were moving rigid areas of high contrast. It did not particularly matter whether we used dots, as in our previous studies, wires, as in the early wire-frame KDE work (Wallach & O'Connell, 1953), disks, or patterned disks. The disk and patterned disk stimuli provided very strong percepts of shape, although the disks did not undergo realistic foreshortening as they rotated. In fact, the dot stimuli gave the weakest percept of depth. These tokens had the least contrast energy (i.e. were the smallest), and hence were harder to detect. Subject JBL had the greatest difficulty in seeing these small dots, and his results show a slight drop in performance for the dot stimuli.

Dynamic-on-gray. The motion of a token filled with dynamic random dot noise moving on a gray background is microbalanced. In other words, 1st-order motion detectors are "blind" to this stimulus. The expected value of the output of such a detector is zero (across random realizations of the stimulus). Simple 2nd-order mechanisms (e.g. using rectification) serve to reveal the true motion.

The results for three subjects are somewhat different. For two subjects (LJJ and JBL), performance is always at or near chance (less than 10% correct in all cases), although for subject LJJ with the dynamic-on-gray dots the performance is significantly above chance ($P < 0.05$). On the other hand, for subject MSL, performance is always well above chance (24–39% correct identifications), but far less than his nearly perfect (94–98% correct) performance with white or pattern tokens on gray.*

The 1st-order motion mechanisms are clearly the most effective input to the KDE system, since eliminating motion detectable by 1st-order mechanisms reduces performance substantially for all subjects. The results for subject MSL suggest that 2nd-order motion mechanisms can also be used. On some trials, fragments of the microbalanced stimuli did appear 3D to this subject (one of the authors), especially in the foveally-viewed portion of the stimulus. To raise his performance level, he used sophisticated guessing strategies based on active eye movements and local measurements of motion or three-dimensionality in the fovea at a small number of locations of the display. But, these strategies only serve to bring performance up to mediocre levels in comparison with performance with rigid white-on-gray motion.

Dynamic-on-static. The dynamic-on-static manipulation also results in a micro-balanced stimulus. For the dynamic-on-static conditions, performance is at chance level for all three subjects, and for both wire disk tokens. As with the dynamic-on-gray conditions, the motion of the tokens is visible. It is not particularly difficult to detect the motion of an area of dynamic noise on a static noise background (Chubb & Sperling, 1988b). However, this sort of motion engenders no shape percept whatever under the conditions of our experiments.

Unlike dynamic-on-gray stimuli, dynamic-on-static stimuli are not revealed by contrast rectification. Detection of the motion of a region of flicker requires more elaborate 2nd-order mechanisms. Regions of flicker could first be detected by applying a linear temporal filter (such as differentiation), followed by rectification, and then by application of a 1st-order motion mechanism. Some such complex 2nd-order motion detector exists in the human visual system, since we are capable of seeing areas of flicker move, including in the displays of our experiment (at least with scrutiny). Yet, this 2nd-order motion detection system does not support the structure-from-motion computation for our dynamic-on-static stimuli.

Prazdny (1986) reached the opposite conclusion using dynamic-on-static displays representing simple wire objects rotating in a tumbling motion. Each object contained five wires, and subjects were required to identify the object among six alternative wire-frame objects.

*In order to test the range of luminances over which polarity alteration was effective, we ran a control experiment (using MSL and JBL as subjects), where a variety of white pixel luminances were used with a given black pixel luminance. We viewed a variety of dynamic-on-gray displays, varying the luminance values for the black and white pixels independently over a wide range. We also tested a variety of other luminance calibration procedures. Dynamic-on-gray stimuli are only micro-balanced if the contrast energy of the white pixels is the same as that of the black pixels. And, it is difficult to calibrate the luminance of individual pixels embedded in a complex display texture given that the desired pattern is first low-pass filtered by the CRT video amplifier, and then passes through the gun nonlinearity (see Mulligan & Stone, 1989, for a full discussion of this point). Thus, it was important to verify that our results were robust over a range of luminance values overlapping the calibrated equal contrast point.

To summarize, shape identification performance is consistent with the results of expt 1 for a reasonably wide range of white pixel luminances. Subject MSL consistently performs at moderate levels, and subject JBL consistently performs at or near chance. The luminance levels yielding poor shape identification performance are consistent with the levels that result in the weakest 3D percept, and are roughly consistent with the luminance levels that are balanced (black pixel decrement vs white pixel increment) for a variety of calibration displays. The performance levels for dynamic-on-gray stimuli in expt 1 do not result from a miscalibration of luminance levels.

The displays were 7 × 7 deg, and the wires were several pixels thick. Performance was quite high in the task for five subjects. Although we have some reservations about the experimental method employed by Prazdny, we have generated similar displays in our laboratory, and our dynamic-on-static wire-frame displays do yield a shape percept when displays are restricted to a small number of wires.

The most likely explanation of the difference between our results and those of Prazdny involves the difference in spatial resolution required by each task. Chubb and Sperling (1988a) have demonstrated that 2nd-order motion systems have less spatial resolution than the 1st-order mechanisms, and that their resolution drops precipitously with increases in retinal eccentricity. In our displays, motion was about a vertical axis using parallel perspective, and hence all motion was along the horizontal. There could be as many as 10 or 20 disks or wires in a given row of the image to resolve. Our displays did not yield a global percept of optic flow, but motion was perceived foveally with scrutiny. This is entirely consistent with Chubb and Sperling's observation. Prazdny did not give precise details about his stimuli, but it was clear that along a given motion path there were only two or three wires to resolve across his far larger display. Performance was so low in our dynamic-on-static conditions because too much spatial acuity was required of the 2nd-order system that detects the motion of flickering regions.

How useful for perception of shape is a display of dynamic noise figures moving on a static noise background? We have examined a large number of disk and (thick) wire displays in order to span the gap of spatial resolution between Prazdny's displays and our own. With our 3 × 3 deg display size, a shape percept can only be achieved by using a very small number of tokens (around 5–10). These displays consisted of rotating disk tokens. Cavanagh and Ramachandran (1988) suggest an alternative explanation of the difference between our results and those of Prazdny. They consider the crucial difference to be that the objects portrayed in the Prazdny displays were connected (one long wire figure), whereas our displays consisted of separate disk tokens. With our wire displays, almost no 3D percept was achieved for the dynamic-on-static condition. In addition, we were able to achieve a 3D percept with displays of a small number of dynamic-on-static disks. Thus, we feel that low spatial resolution in the 2nd-order motion system (rather than unconnected tokens) is the likely explanation for failure of KDE.

Contrast polarity alternation. Performance is quite poor for the contrast polarity-alternating dots as it was in the previous paper (Dosher et al., 1989b). For two subjects (JBL and LJJ) performance is at chance or insignificantly above chance. For subject MSL, performance is low (11% correct) but significantly above chance ($P < 0.05$). On the other hand, when the token is changed to disks or wires, performance rises substantially. Contrast polarity alternation is not as devastating a stimulus manipulation for disks and wires as it is for dots.

For 1st-order motion detection mechanisms such as the Reichardt detector, contrast polarity alternation causes the strongest responses to be in the wrong direction. Yet, the intended motion can be detected quite accurately if a 2nd-order detector is used that first applies a luminance nonlinearity followed by a Reichardt detector. The primary difference between the dots on the one hand, and the disks and wires on the other, is that the disks and wires have more pixels illuminated. In other words, they have more contrast energy, and in particular thay have more energy at lower spatial frequencies. Thus, the disk and wire stimuli should stimulate both the 1st- and 2nd-order motion detection systems more strongly, resulting in stronger incorrect direction information from the 1st-order system as a whole, but also stronger information from the 2nd-order system, and stronger directional information in those selected 1st-order frequency bands which signal the correct direction.

It is interesting to note that a large number of the errors made by observers with polarity-alternating stimuli were errors in the direction of rotation *only*, with the shape specified correctly. For example, for a stimulus which had as correct answers either $u + - 0l$ or $u - + 0r$, the subject incorrectly responded with $u + - 0r$ or $u - + 0l$, rather than with any of the 104 other possible incorrect responses. This effect was largest for the disk tokens. In a separate control experiment, for contrast polarity-alternating disk stimuli, 39% of the errors made by subject MSL were only an error in the specification of direction, compared to 1.4% direction errors for the dynamic-on-gray conditions. For subject JBL, the corresponding values were 48% and 5.6%. For the polarity-alternating disks, on trials when subject MSL correctly identified the shape, there was a 33% chance that he would misidentify the direction of rotation (for JBL: 29.3%). We believe that accurate shape identification in this condition primarily reflects responses constructed from selected 1st-order information. One strategy was simply to specify the opposite rotation direction to that which was perceived! The displays did, however, occasionally appear to be 3D with the correct direction of motion (at certain times during the rotation, or close to the location to which the eyes were directed), indicating a residual 2nd-order motion input to the KDE system. The fact that these displays only appeared foveally to be rotating in the correct direction, and then only using the larger tokens, is consistent with a 2nd-order motion detection system with low contrast sensitivity and low spatial resolution (as has been demonstrated by Chubb & Sperling, 1988b), and more sensitive in the fovea (Chubb & Sperling, 1988a). In summary, we have some indication that 2nd-order motion detection mechanisms can be used to derive 3D structure, but they are far less robust and have poorer spatial resolution than 1st-order motion mechanisms.

Pattern-on-static. For all three subjects performance with pattern-on-static displays is quite poor (9, 26 and 33% corrrect), although it is significantly above chance levels in all cases ($P < 0.05$). This poor performance results from a mismatch of resolution and temporal sampling. The patterned disks are quite detailed/high frequency. The disks are 6 pixels in diameter, and can move as far as 8.3 pixels in one frame. This speed is only achieved by disks at the top of a peak when in the middle of the display (i.e. near frame numbers 0, 15 and 29), but many disks are moving 3–5 pixels per frame. High frequency spatial filters which are required to identify the disks must correlate across frames with filters that are far more than 90 deg away in the phase of their peak spatial frequency. A typical 1st-order detector will not compare spatial regions that far apart in order to avoid spatio-temporal aliasing (van Santen & Sperling, 1984). Thus, the clearest motion signals are coming from the slower areas in the display, which are the least useful for discriminating the shapes. We have examined pattern-on-static displays with finer temporal sampling (60 new frames per sec, as opposed to 4 repaints of 15 new frames per sec used in the experiment), and they give a strong impression of three-dimensionality. Thus, poor performance in the task resulted from undersampling in time of the stimuli, which interferes with 1st-order (and some 2nd-order) motion mechanisms, and good KDE can result from the motion of tokens which are camouflaged when at rest.

We have also examined dynamic-on-static displays with finer temporal sampling (60 new frames per sec). These displays yield no impression of three-dimensionality. The poor results for dynamic-on-static displays do not result from insufficient sampling in time. Also, since finely sampled pattern-on-static displays do appear 3D, poor performance with dynamic-on-static-displays does not result from the camouflage of the tokens when at rest. Rather, dynamic-on-static displays yield no effective KDE because of the low resolution of the 2nd-order system required to analyze the motion.

EXPERIMENT 2. TWO-FRAME KDE

The first experiment shows that accurate performance in shape identification is dependent upon a global (primarily 1st-order) optic flow. If a stimulus manipulation makes that optic flow noisy or otherwise interferes with the optic flow computation, there is little or no KDE. This occurs even though foveal scrutiny does reveal the motion in these displays.

If the percept of surface shape depends upon a global optic flow, then we should be able to get reasonable shape identification performance from any stimulus that results in a strong percept of optic flow. In particular, the extended (2 sec) viewing conditions of expt 1 should not be necessary. Two frames are obviously the minimum number of frames that can yield a percept of motion, and two frames should suffice. In the second experiment, we investigate the accuracy of performance in the shape identification task for two-frame displays.

Method

Subjects. There were two subjects in this experiment. One was an author, and the other was a graduate student naive to the purposes of this experiment. Both had normal or corrected-to-normal vision. There were slight differences in the conditions for each of the two subjects. These will be pointed out below.

Stimuli and apparatus. The stimuli were similar to the white-on-gray dot stimuli from expt 1. Stimuli were generated from the same set of 3D shapes, using the same dot densities, and projected in the same way. The local dot density was kept constant using the same scintillation procedure. New stimuli were computed, two of the flat shape, and one of each of the other 52 shapes, resulting in 54 displays.

Each display consisted of 11 frames, rotating from 20 deg left to 20 deg right in increments of 4 deg per frame. The middle frame (number 6) was face-forward, as was the first frame of each display in expt 1. Two-frame stimuli consisted of a presentation of the middle frame followed by one of the other 10 display frames. This resulted in either a leftward or rightward rotation of 4–20 deg between the two frames of the display. A single trial display consisted of 0.5 sec of a cue spot, 0.5 sec blank, the first frame, an inter-stimulus blank interval (or ISI), the second frame, and a blank. Each stimulus frame was repainted four times at 60 Hz, for a total duration of 67 msec. We define the ISI to be the time interval between the onset of the last painting of the first stimulus frame and the onset of the first painting of the second stimulus frame. For example, when no blank frames were used, the ISI was 16.7 msec. Displays were 182 × 182 pixels, and were presented using the same apparatus and viewing conditions as for subject LJJ in expt 1. The background luminances for subjects MSL and LJJ were 15.6 cd/m^2 and 5.0 cd/m^2, respectively. The corresponding dot luminosities were 26.8 and 15.7 extra μcd/dot, respectively. Nominal contrasts were huge (i.e. nominal Weber contrasts of 500% or more).

Procedure. The task was shape and rotation identification. Subjects keyed their responses using response buttons, and received feedback on the display after their response. Three groups of trials were run. In the first, the ISI was 16.7 msec, and rotation angle between frames was varied from 4 to 20 deg. Since the second frame could be chosen from either the frames preceding or succeeding the middle frame (rotation to the left or right), this resulted in 540 possible stimuli (54 displays, 2 directions, 5 rotation angles). These were run in random order in 4 blocks of 135 trials. In the second group of trials, rotation was kept constant at 4 deg. ISI ranged from 16.7 to 83.3 msec. This again resulted in 540 trials presented in random order in 4 blocks of 135 trials. In the third group

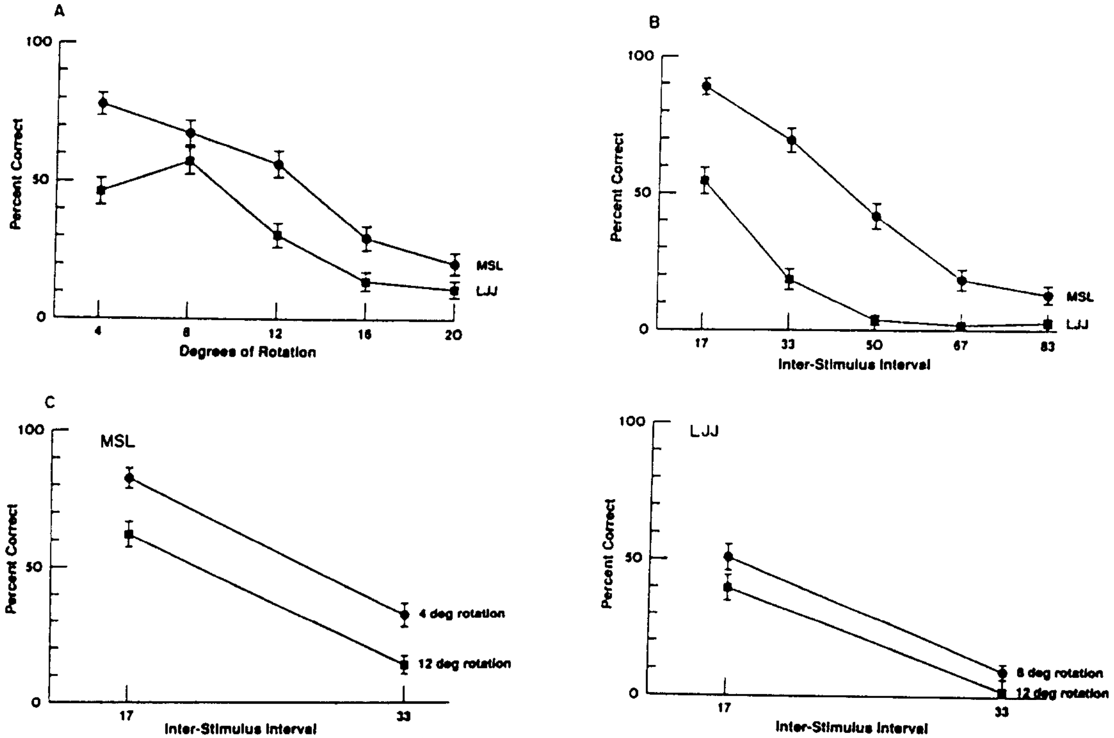

Fig. 4. Results of expt 2. Data for two subjects are shown. Error bars indicate ±1 SEM. (A) Shape-and-rotation identification accuracy as a function of the angle of rotation between the two frames. ISI was 16.7 msec. (B) Shape-and-rotation identification accuracy as a function of the duration of a blank inter-stimulus interval (ISI). Rotation angle was 4 deg. (C) The two manipulations used in the same experiment. Note the lack of interaction.

of trials, both rotation angle and ISI were varied. The ISIs were either 16.7 or 33.3 msec. For subject MSL, the rotation angles were either 4 or 12 deg. For LJJ, they were either 8 or 12 deg. These four conditions (two rotation angles by two ISIs) resulted in 432 trials which were presented in random order in 4 blocks of 108 trials.

Results

The results are shown in Fig. 4. Each data point is the percent correct over 108 trials. As is evident from the figure, shape identification can be quite high for these minimal motion displays (for similar observations using different experimental methodology, see Braunstein, Hoffman, Shapiro, Andersen & Bennett, 1987; Lappin, Doner & Kottas, 1980; Mather, 1989; and Petersik, 1980). For an ISI of 16.7 msec (Fig. 4A), this entire sequence lasted only 133 msec. Yet, performance was as high as 54.6% for subject LJJ, and 88.9% for subject MSL (62.8% and 94.2% of their white-on-gray dots performance in expt 1, respectively). Two frames of moving dots are sufficient for accurate, although not perfect

performance in this shape identification task. Since these experiments were first reported (Landy, Sperling, Dosher & Perkins, 1987a; Landy, Sperling, Perkins & Dosher, 1987b), Todd (1988) has also shown above-chance KDE performance for two-frame stimuli, although in his paradigm the two frames are repeated several times before a response is made.

Rotation angle and fixation. Performance as a function of rotation angle between the two frames is given in Fig. 4A. Performance decreases with increasing angle of rotation for subject MSL. For subject LJJ, performance reaches a peak at 8 deg, and decreases for smaller and larger rotations. The decrease in performance with larger rotation angles is to be expected, since the correspondence problem becomes increasingly difficult as dots move farther from their initial positions. One might also expect performance to drop as rotation angle decreases to zero. At extremely small rotation angles, the remaining motion would fall below threshold. In our displays, the drop with small rotation angles might be expected to occur even sooner as the small motions in the display became corrupted by poor spatial sampling (inter-pixel distance was approx. 1 min arc). This drop was only seen in the data of LJJ, and presumably would be seen in those of MSL if he had been tested using smaller rotations.

In a previous paper (Dosher et al., 1989b), we found that adding a blank interval between successive frames of a 30 frame KDE stimulus reduced shape identification to near chance performance. This was explained by reduction of power in the stimulus to the 1st-order system. This effect is also seen here, where performance decreases monotonically with increasing ISI (Fig. 4B). Subject LJJ performs at chance levels with a 50 msec or greater ISI, while subject MSL is still slightly above chance performance with an 83.3 msec ISI.

Time and distance. In the previous two groups of trials, there was a confounding between the stimulus manipulation (rotation angle or ISI) and dot velocity. Greater rotation angles at a fixed (16.7 msec) ISI produced greater velocities. Similarly, greater ISIs at a fixed 4 deg rotation angle resulted in smaller velocities. If performance were simply a function of velocity, then rotation angle and ISI should trade off. In Fig. 4C we present the results of varying both ISI and rotation angle factorially. We used a different set of rotations for subject LJJ than MSL based on the results in Fig. 4A, so that for both subjects the performance was expected to decrease with increasing rotation angles. As can be seen in the figure, the two variables do not trade off as would be expected if performance were only a function of velocity, or rotation speed. Increasing rotation angle increases the difficulty of the correspondence problem. Increasing ISI causes increasing problems for the motion detection system. Both manipulations degrade performance in an additive fashion. This observation contradicts Korte's (1915) 3rd law of apparent motion perception, which states that an increase in ISI must be counteracted by an increase in distance traveled for strong apparent motion. In Fig. 4C, Korte's law predicts a cross-over interaction, which is strongly disconfirmed. However, Burt and Sperling (1981) show that time and distance have independent additive effects on the strength of the apparent motion of dot stimuli, which agrees with the present results.

KDE from optic flow. Accurate KDE performance requires a global optic flow. When that optic flow is produced by a minimal motion stimulus—a two-frame display—the shape percept may be fragile and easily degraded by a variety of stimulus manipulations. The stimuli are quite brief in this paradigm and, by subject reports, appear as a collection of dots moving at various speeds, i.e. "look like" an optic flow. On some trials, only patches of planar motion are perceived, and the shape response is generated cognitively. On other trials, a 3D surface is perceived. On some trials the optic flow is perceived and so is the shape, but the shape percept is only "felt" after the display is over. As we discussed extensively in our first article on the shape identification task (Sperling et al., 1989), KDE is inextricably tied with the percept of an optic flow. It can be very difficult to differentiate empirically between a judgment based on a 3D percept and performance based on an alternative strategy (computationally equivalent to that required for KDE) using a remembered set of 2D velocities.

Reasonably accurate performance on the shape-and-rotation identification task results from only two frames of 300 points. In the computer vision literature, there have been several studies of the structure-from-motion problem resulting in theorems of the following form: "*m* views of *n* points under the following restrictions of the motion path suffice to determine the 3D structure up to a reflection" (Bennett & Hoffman, 1985; Hoffman & Bennett, 1985; Hoffman & Flinchbaugh, 1982; Ullman, 1979). It has been suggested that these minimal conditions for structure from motion also govern human perception (Braunstein et al., 1987; Petersik, 1987). The particular models just mentioned do not have any prediction concerning performance in the 300 points/2 views situation used here. An exception is a recent paper by Bennett, Hoffman, Nicola and Prakash (1989), where it is shown that there is a one parameter family of possible interpretations for two frames of four or more points. This family is parameterized by the slant of the axis of rotation (as in the "isokinescopic displays" described by Adelson, 1985), and the paper does not deal explicitly with rotation axes in the image plane, as used here. On the other hand, models that compute 3D structure based only upon a single velocity field do allow for this performance (Longuet-Higgins & Prazdny, 1980; Koenderink & van Doorn, 1986). We take our experimental results as evidence for optic flow-based methods for the KDE, as opposed to models requiring three or more views. In particular, our results strongly rule out models that require measurement of acceleration in addition to velocity (e.g. Hoffman, 1982).

Structure-from-motion computation may improve its 3D representation with additional information (e.g. with additional frames, Grzywacz, Hildreth, Inada & Adelson, 1988; Hildreth & Grzywacz, 1986; Landy, 1987; Ullman, 1984). The shape in our two-frame displays does not always appear to have the depth extent that results from the 30 frame displays of expt 1, and two-frame performance is reduced relative to 30-frame performance. The shape identification task can be solved by knowing only the sign of depth and direction of motion in each spatial location (up to a reflection), without accurately estimating either velocity or the amount of depth.

DISCUSSION

Two experiments investigated the type of motion detection mechanism used as an input to the structure-from-motion system. Performance in the shape-and-rotation identification task was accurate regardless of the token used to carry the motion, as long as that token was presented with constant contrast polarity (the white-on-gray and pattern-on-gray conditions). The performance decrements seen with contrast polarity alternation and the two microbalanced conditions add further evidence to the conclusion of Dosher et al. (1989b) that 1st-order motion detectors are the primary substrate for the computation of shape. In addition, there are indications of an input to the shape computation from 2nd-order motion mechanisms, which is weak, low in spatial resolution, and concentrated at the fovea. 2nd-order mechanisms that require temporal filtering (i.e. detection of flicker) prior to a point nonlinearity were useless here because of the spatial resolution required by our stimuli. These sorts of detectors would only be useful for KDE displays involving a small number of moving features, rather than the densely sampled optic flows required for the determination of precise shapes of curved surfaces from motion cues. The results from the two-frame experiments reinforced these conclusions. They also demonstrated that detection of instantaneous velocity is sufficient for KDE; acceleration is not required, nor are more than two views.

Acknowledgements—The work described in this paper was supported primarily by a grant from the Office of Naval Research, grant N00014-85-K-0077, and partly by USAF Life Science Directorate, grants 85-0364, 88-0140, and NSF grant IST-8418867. We would like to thank Charles Chubb for his helpful comments, and Robert Picardi for technical assistance. Portions of this work have been presented at the annual meetings of the Association for Research on Vision and Ophthalmology, Sarasota, Florida (Landy et al., 1987a) and the Optical Society of America, Rochester, New York (Landy et al., 1987b).

REFERENCES

Adelson, E. H. (1985). Rigid objects appear highly nonrigid. *Investigative Ophthalmology and Visual Science* (Suppl.), *26*, 56.

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. & Rogers, B. J. (1975). Illusory reversal of depth and movement during changes of contrast. *Vision Research*, *15*, 957–961.

Bennett, B. M. & Hoffman, D. D. (1985). The computation of structure from fixed-axis motion: Nonrigid structures. *Biological Cybernetics*, *51*, 293–300.

Bennett, B. M., Hoffman, D. D., Nicola, J. E. & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052–1069.

Braunstein, M. L., Hoffman D. D., Shapiro, L. R., Andersen, G. J. & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 335–343.

Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London, B*, *227*, 249–265.

Burt, P. & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, *88*, 171–195.

Cavanagh, P. & Ramachandran, V. S. (1988). Structure from motion with equiluminous stimuli. Paper presented to the *Annual Meeting of the Canadian Psychological Association*, Montreal, June.

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

Chubb, C. & Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, *5*, 1986–2007.

Chubb, C. & Sperling, G. (1989a). Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, U.S.A.*, *86*, 2985–2989.

Chubb, C. & Sperling, G. (1989b). Second-order motion perception: Space/time separable mechanisms. *Proceedings: Workshop on visual motion* (pp. 126–138). Washington, D.C.: IEEE Computer Society Press.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989a). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 816–825.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989b). The kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research*, *29*, 1789–1813.

Grzywacz, N. M., Hildreth, E. C., Inada, V. K. & Adelson, E. H. (1988). The temporal integration of 3-D structure from motion: A computational and psychophysical study. In von Seelen, W., Shaw, G. & Leinhos, U. M. (Eds.), *Organization of neural networks*. New York: VCH.

Heeger, G. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A*, *4*, 1455–1471.

Hildreth, E. C. & Grzywacz, N. M. (1986). The incremental recovery of structure from motion: Position vs velocity based formulations. *Proceedings of the workshop on motion: Representation and analysis*. IEEE Computer Society no. 696, Charleston, South Carolina, 7–9 May.

Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America 72*, 888–892.

Hoffman, D. D. & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, *2*, 350–353.

Hoffman D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, *42*, 195–204.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: The University of Chicago Press.

Koenderink, J. J. & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A*, *3*, 242–249.

Korte, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, *72*, 193–206.

Landy, M. S. (1987). A parallel model of the kinetic depth effect using local computations. *Journal of the Optical Society of America A*, *4*, 864–876.

Landy, M. S., Cohen, Y. & Sperling, G. (1984a). HIPS: A Unix-based image processing system. *Computer Vision, Graphics and Image Processing*, *25*, 331–347.

Landy, M. S., Cohen, Y. & Sperling, G. (1984b). HIPS: Image processing under UNIX—Software and applications. *Behavior Research Methods, Instruments and Computers*, *16*, 199–216.

Landy, M. S., Sperling, G., Dosher, B. A. & Perkins, M. E. (1987a). Structure from what kinds of motion? *Investigative Ophthalmology and Visual Science* (*Suppl.*), *28*, 233.

Landy, M. S., Sperling, G., Perkins, M. E. & Dosher, B. A. (1987b). Perception of complex shape from optic flow. *Journal of the Optical Society of America A*, *4*, 108.

Lappin, J. S., Doner, J. F. & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, *209*, 717–719.

Lelkens, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual display. *Vision Research*, *24*, 1083–1090.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, *208*, 385–397.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, *211*, 151–180.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *The Quarterly Journal of Experimental Psychology*, *41A*, 183–198.

Mulligan, J. B. & Stone, L. S. (1989). Halftoning method for the generation of motion stimuli. *Journal of the Optical Society of America A*, *6*, 1217–1227.

Petersik, J. T. (1980). The effects of spatial and temporal factors on the perception of stoboscopic rotation simulations. *Perception*, *9*, 271–283.

Petersik, J. T. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception and Psychophysics, 42*, 355–364.

Prazdny, K. (1986). Three-dimensional structure from long-range apparent motion. *Perception, 15*, 619–625.

Ramachandran, V. S., Rao, V. M. & Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13*, 1399–1401.

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift Naturforschung B, 12*, 447–457.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*, 451–473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–321.

Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instruments, 3*, 148–151.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8*, 144–151.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). The kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 826–840.

Sperling, G., Dosher, B. A. & Landy, M. S. (1990). How to study the kinetic depth effect experimentally. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 445–450.

Todd, J. T. (1988). Perceived 3D structure from 2-frame apparent motion. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *29*, 265.

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

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception, 13*, 255–274.

Wallach, H. & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Watson, A. B. & Ahumada, A. J. Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 1*, 322–342.