Pergamon

0042-6989(95)00295-2

Vision Res., Vol. 36, No. 15, pp. 2271–2281, 1996
Copyright © 1996 Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0042-6989/96 $15.00 + 0.00

The Detection of Stimuli Rotating in Depth Amid Linear Motion and Rotating Distractors

J. TIMOTHY PETERSIK*

Received 14 December 1994; in revised form 4 August 1995; in final form 23 October 1995

In three experiments, observers watched displays consisting of two or more areas that contained unidirectionally moving pixels. In half of the displays, one area of pixels contained movement that corresponded to the projection of the front surface of a rotating cylinder. The total duration of the displays and the number of stimulus areas per display were varied. The subjects' task was to indicate whether or not a given display contained rotation. When the display time required to reach 75% accuracy was determined, it was found that the number of stimuli per display had no effect; nor did it interact with other variables. One control experiment eliminated "pixel crowding" at the edges of the rotating cylinders, with little effect on the results. Another control experiment found that the ability to discriminate rotating from linear motion declines with distance away from fixation. A fourth experiment showed that under conditions similar to the first three, subjects can make accurate shape discriminations, thereby suggesting that three-dimensional information contributed to the decisions made in the original experiments. On the basis of these results and previous data, it is suggested that in the present experiments structure was recovered from motion by the short-range process, and that this recovery engages attention to a relatively constant extent, regardless of the number of stimuli contained in a display. Shape discrimination based on structure from motion may require a more effortful form of attention. Copyright © 1996 Elsevier Science Ltd.

Motion perception   Structure from motion   Visual attention   Short-range process

INTRODUCTION

Dick *et al.* (1991) examined conditions that facilitated the detection of a three-dimensional rotating stimulus. A rotating cylinder was simulated by pixels that moved in an orthographic projection; both front and back of the cylinder were visible, and hence there were pixel motions both to the right and the left. While the pixels of the simulated cylinder moved, there was also a background of pixels, each of which moved linearly either to the right or to the left with the same average velocity as the target's pixels. Across several experiments, Dick *et al.* (1991) found that detection of the rotation amidst noise motion was very high when the two-dimensional (2D) displacements of its pixels were within the spatial displacement limit of the short-range process (SRP), and that detection declined rapidly as more pixel displacements entered the range of the long-range process (LRP). Because the same authors had previously shown the SRP to be pre-attentive (i.e., reaction time to short-range motion did not increase with the addition of distractors; Dick *et al.*, 1987), their work makes the implicit case that the detection of rotation, under optimal displacement conditions, is also pre-attentive.

The present experiments were conducted in order to further understand the attentional requirements involved in the detection of rotation produced by short-range motion. Like Dick *et al.* (1991), we asked subjects to detect the presence of a rotating stimulus in displays that contained linearly moving pixels. However, a number of changes were made in the stimuli: first, whereas Dick *et al.* (1991) embedded their rotating cylinder in a relatively homogeneous background of linearly moving pixels, in the current displays there were discrete areas occupied by either linearly moving or rotating pixels. This made it possible to examine the effect of set size (e.g., Palmer, 1994), rather than pixel numerosity, on rotation detection. Second, whereas the previous authors varied angular velocity of the rotating stimulus in an effort to manipulate short-range/long-range processing, in the present experiments angular velocity was held constant (along with the average, short-range, displacement of pixels), and the absolute rotation (total duration) of the displays was varied. This permitted us to estimate the time required for subjects to reach a criterion level of decision-making. Third, whereas in the earlier study both left and right 2D motions of the pixels were visible, the present displays

*Department of Psychology, Ripon College, P. O. Box 248, Ripon, WI 54971, U.S.A. [*Email* petersikt@acad.ripon.edu].

were unidirectional, showing either leftward or rightward motion only in a given display (only the "front" surface of a rotating cylinder was displayed and linear noise motion was in the same direction). This separated any potential directional artifacts from pure rotation information. In separate experiments, we also varied the eccentricity of the stimuli and examined the role of "edge-crowding" in rotation detection. Finally, to determine the extent to which subjects actually use 3D structure from motion to make rapid discriminations, we embedded rotating targets (spheres) amidst rotating distractors (cylinders).

Along with the work of Dick *et al.* (1991) suggesting that the most efficient detection of rotation is conducted by a "parallel" process (i.e., the SRP), Shulman (1991) has shown that rotation aftereffects (Petersik *et al.*, 1984) can be modulated by selective attention to specific rotating adaptation stimuli. The former finding implies that rotation detection can be conducted by a rapid, low-level, relatively automatic process (i.e., the SRP); the latter suggests that following detection, the perception of rotation can be, or is, maintained by a higher order, relatively effortful process. This could explain why Dick *et al.* (1991) were able to obtain some rotation detection under LRP conditions, although it was not very efficient. Currently, there is some issue in the literature as to whether visual attention is best considered in terms of a serial/parallel distinction (e.g., Treisman & Gelade, 1980), a pre-attentive vs attentive distinction (e.g., Julesz, 1990), a decision integration approach (e.g., Palmer, 1994), or by some not specifically attentional influence like "discriminability" (e.g., Verghese & Nakayama, 1994). Because the present experiments were designed to better understand the processes underlying the detection of rotating stimuli defined by the short-range motion of pixels, and not as tests of any specific model of visual attention, there is an attempt to remain as theoretically neutral as possible with respect to theories of attention.

The goal of the present experiments was to determine the influence of the stimulus set size on correct rotation detection, and to determine whether set size influences the time required to make a decision about the presence or absence of rotation. The influence of retinal eccentricity, the importance of edges, and the recovery of structure from brief motion in the detection of rotation are considered in separate control experiments.

GENERAL METHODS

Stimuli and apparatus

General construction of stimuli. All stimuli were prepared on an Amiga 600 microcomputer. The overall strategy was to prepare small area "micro-displays" of collections of pixels specifying the rotation of the near surface of a cylinder (i.e., the surface facing the observer), along with micro-displays showing the same set of pixels in linear motion with the same average velocity. First, 11 pairs of pixels (i.e., pixels adjacent to one another in either the vertical or horizontal dimension, randomly determined) were randomly positioned in a small area of the computer screen. From these, rotating stimuli were prepared using the techniques described in Petersik (1991b); i.e., the positions of pixel pairs in 29 subsequent frames of the display were determined by conventional methods (e.g., Braunstein, 1976), and all frames were stored to create a 30 frame micro-display showing the pixels rotating through 180 deg (thereby producing a rotational velocity of 6 deg per frame). Linear motion micro-displays were prepared by using the same initial collection of random pixel pairs in Frame 1. The final horizontal locations of the pixel pairs in the rotation micro-displays was also determined. For the linear motion micro-displays, the pixel-pairs were subsequently displaced in equal steps across the next 29 frames so as to arrive at the same horizontal locations as their counterparts in the rotating micro-displays. In both cases, the disappearance of a pixel pair at one edge of a display occasioned the appearance of a new pixel pair randomly located (but the same in both types of micro-displays) at the opposite edge in the subsequent frame of the display.

For experimental conditions that required fewer than 30 frames per micro-display, the unnecessary frames were deleted from the beginning or end of the original, 30 frame, display. Thus, all micro-displays maintained the same rotational or linear velocity and the spatial arrangement of pixels; only their total duration and the absolute distance traversed by individual pixel pairs varied from display to display.

Micro-displays were stored on hard disk. They could subsequently be positioned anywhere on the computer screen in the preparation of "macro-displays". Macro-displays were animations whose components were the micro-displays described above. The resulting macro-displays thereby showed a variable number of collections of pixel pairs, each of which could either display rotation or linear motion. With the exception of Experiment 4, no more than one rotating micro-display was ever used in a macro-display. For the main experiment described here, pixel pairs in both rotation and linear motion micro-displays always moved from left to right. For the rotation direction control experiment described in the Results section, motion direction was randomly determined; direction could be varied by presenting the micro-displays in either a forward or backward order of frames.

Details of stimuli.. Viewing distance was set at 66 cm. The 200 pixel (vertical) × 320 pixel (horizontal) display area of the monitor screen (Commodore-Amiga, model 1084S) thereby subtended 13.5 deg × 21 deg visual angle. Each micro-display was created within an approximately square area subtending 2.78 deg (vertical) × 3.04 deg (horizontal); i.e., 41 pixels × 46 pixels. The background of the display area, as well as the background of each micro-display, was kept dark (0.8 cd/m^2). Each pixel in a display was white (25 cd/m^2). The average horizontal distance traversed between frames by pixels was 11.4′ visual angle, well within the putative spatial limit of the SRP of 15′ visual angle for small stimuli (cf. Petersik *et*

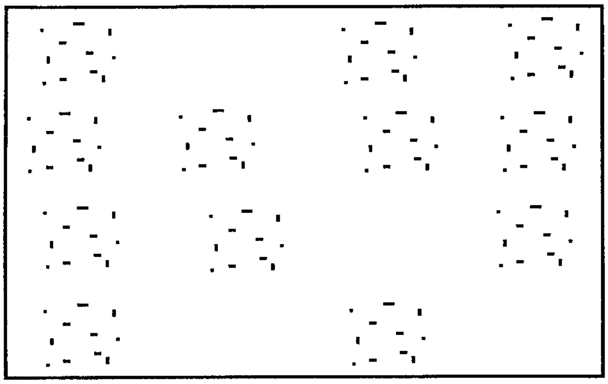

FIGURE 1. Computer-generated drawing of the general appearance of one frame of a 12 stimulus display of the kind used in the present experiments. Note that contrast is reversed in this picture and that pixels appear larger than they were in the actual experiments.

al., 1983). Our measurements of the displacements of pixels in the rotation simulations showed that four pixels made displacements greater than 15′ visual angle, the greatest of these being 19′; we conclude therefore that most of the rotation information contained in these displays was limited to the spatial range of the SRP.

Micro-displays that simulated the rotation of a cylinder around its longitudinal (i.e., vertical) axis were prepared in a polar projection with a perspective ratio (simulated viewing distance divided by cylinder radius) of 3.0.

Macro-displays were created that consisted of 5, 10, 15, and 30 frames, each constructed from the parent displays described above. When macro-displays were constructed, micro-displays were placed randomly within the cells of an imaginary 4 × 4 grid on the monitor screen (except in Experiment 2); each cell subtended approximately 3.38 deg (vertical) × 5.25 deg (horizontal). Each micro-display was confined to this area, but did not necessarily occupy the exact center. Three different sets of micro-displays, and therefore macro-displays, were constructed in order to establish a population of stimuli from which to sample for the subsequent experiments. Figure 1 shows diagrammatically what a single frame of these displays looked like.

Stimuli were presented with the monitor operating in a non-interlaced mode. Frame duration was 1/60 sec; therefore, macro-displays consisting of only five frames of movement lasted slightly over 83 msec, while displays consisting of 15 frames of movement lasted 250 msec. Only macro-displays containing 30 frames of movement, lasting 500 msec, could have been expected to elicit eye movements that would reliably lead to fixations of target micro-displays before their disappearance. Table 1 shows the relationships between the variable number of frames used in displays, the duration of the movement in the displays, and the absolute angular rotation of the cylinder simulations. Referral to this table will assist in the interpretation of data shown in later sections.

TABLE 1. Relationships between number of frames per display, duration of subsequent movement and absolute angular rotation

	Number of frames per display			
Parameter	5	10	15	30
Duration of movement (msec)	83.35	166.67	250.01	500.01
Absolute angular rotation (deg)	30	60	90	180

EXPERIMENT 1: EFFECTS OF SET SIZE AND NUMBER OF FRAMES

Subjects

Subjects consisted of two paid assistants, the author, and an unpaid volunteer. The assistants and volunteer were female, aged 19–21 yr. The author was male, aged 41 yr. There was no visible difference in the data as a function of age or gender. All subjects reported 20/20 vision and good depth perception, either with or without corrective lenses. When corrective lenses were indicated, they were worn throughout testing.

Stimuli and procedure

Micro-displays were grouped randomly within the 4 × 4 grid described above. There were three factors to the experiment: number of frames (that displayed motion): 5, 10, 15, or 30; number of stimuli (or micro-displays) per display: 2, 6, 9, 12, or 16; and presence or absence of rotation. Stimuli were factorially combined and randomly drawn from the larger populations described in the General Methods section. The 40 possible conditions (4 number of frames × 5 number of stimuli × 2 presence/absence) were run in blocks of trials 20 times for each subject. Within each block of trials, the stimuli were presented in a randomized order. A single block of trials was run in a single experimental session, successive sessions typically separated by no less than 24 hr, but occasionally by no less than 1 hr.

For each trial, the subject's task was to stare at the center of the monitor screen, which was dark and blank for 250 msec, and to maintain that fixation throughout the trial. Because of the grid-like arrangement of the stimuli, no stimulus was ever presented directly in fixation. Pixels in motion appeared abruptly and ended with the screen going dark and blank, at which time the subject was to say "yes" or "no", indicating whether or not rotation had been detected.

Results and discussion

Using the percentage of correct responses per condition as the dependent variable, a 2 (rotation present vs rotation absent) × 4 (number of frames per display) × 5 (number of stimuli per frame) repeated measures analysis of variance was conducted. This analysis showed that there was no difference in the percentage of correct responses as a function of whether a rotating stimulus was or was not present in the display, $F(1,3) = 5.72$, $P > 0.05$. Therefore, the stimulus present vs absent factor was not considered in any further analyses.

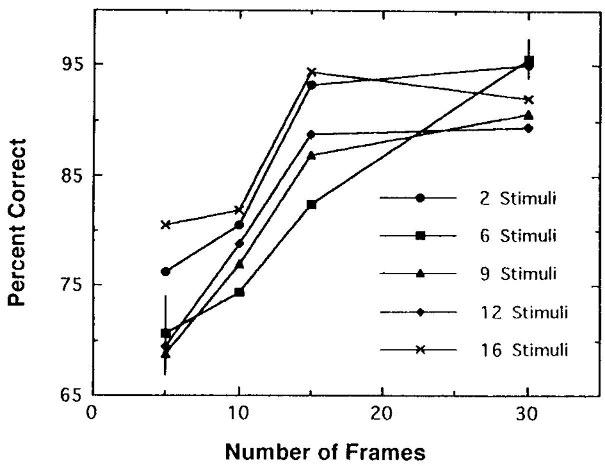

FIGURE 2. Results of Experiment 1, expressed as the percentage of correct rotation detection judgments as a function of the number of frames per display (each frame lasted 16.67 msec). Number of stimuli, or micro-displays, per display is the parameter. Error bars show largest and smallest ±1 SE.

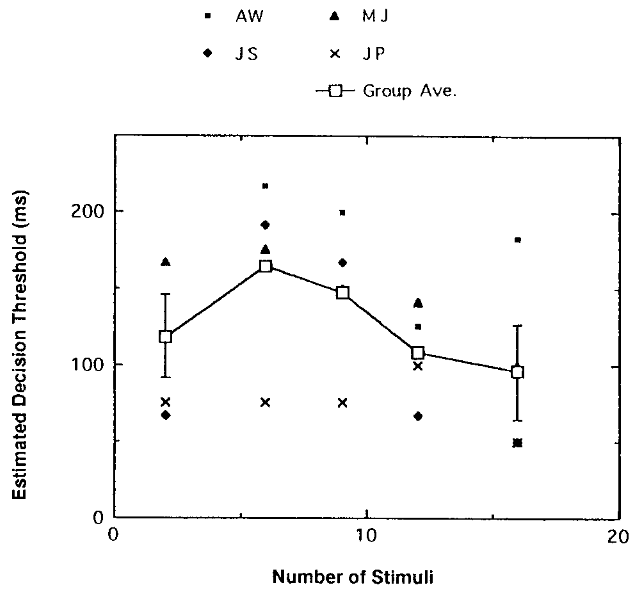

FIGURE 3. Results of Experiment 1, expressed as the 75% decision threshold (msec) as a function of the number of stimuli per display. Data are shown for individual subjects, along with the group average. Error bars show typical standard deviations ±1 SD.

The overall percentages of correct responses as a function of the number of frames per display are shown in Fig. 2; the number of stimuli in each frame of the display is the parameter. The smallest and largest standard errors corresponding to the means are also shown in Fig. 2; the smallest standard error (SE; 2.48%) occurred in the condition that contained six micro-displays over 30 frames, whereas the largest (10.2%) occurred in the condition that contained nine micro-displays over five frames. As can be seen from the means, it was generally the case that the more frames contained in a display, the greater was the percentage of correct responses. This effect was significant in the analysis of variance, $F(3,9) = 29.07$, $P < 0.05$. However, the relationship between the number of stimuli contained in a display and the percentage of correct responses was non-monotonic: subjects were most accurate when displays contained either 2 or 16 stimuli, and were somewhat less accurate when the displays contained six, nine, or twelve stimuli. The effect of the number of stimuli per display was also significant, $F(4,12) = 3.38$, $P < 0.05$. The interaction between the "number of stimuli" and "number of frames" conditions was non-significant, $F(12,36) = 1.50$, $P > 0.05$; thus, the effects of those two factors appear to have been independent and additive.

Whereas subjects were significantly influenced by the number of stimuli contained in the displays, it was clear that the relationship was not so simple as to conclude that the addition of stimuli to a display increased the processing load required of the subjects in a proportional manner. In order to examine more specifically the relationship between the number of stimuli and processing time, we examined each subject's data separately and used linear interpolation to determine the overall display time required to achieve 75% accuracy (referred to as the decision threshold) as a function of the number of stimuli contained in a display. In most cases, this procedure was straightforward. However, of the 20 functions considered, there was one that showed nonmonotonicity; i.e., the function crossed the 75% point twice. In this case, the second crossover was used to estimate the decision threshold. Also, there were three cases in which the functions never fell below 75% correct; in these cases, we found the midpoint between the percentage correct obtained with the 5 frame movies and 0% (for 0 frame movies). The resulting data for each of the subjects are shown in Fig. 3. The shortest 75% decision thresholds were obtained for displays containing 16, 12, and 2 stimuli, respectively. Comparing the 16 and 2 stimuli displays, the short decision thresholds suggest that for these subjects there was no trade-off between time and accuracy in this experiment (i.e., subjects required roughly the same amount of time to achieve 75% accuracy for these displays). Using the logic of Treisman & Gelade (1980), we sought to determine whether there was *any* significant change in the 75% decision threshold as a function of the number of stimuli in a display. A repeated measures analysis of variance on the data shown in Fig. 3 revealed that 75% decision thresholds did not change significantly with the number of stimuli per display $F(4,12) = 1.34$, $P > 0.05$. Therefore, it was tentatively assumed that either (a) the detection of rotating stimuli engages a relatively effortless, low-level attentive process; or (b) the processing load engaged by rotating stimuli is relatively constant and does not fluctuate greatly as non-target stimuli are added to the background.

The macro-displays used in Experiment 1 consisted of sets of initially identically positioned pixels, all of which traversed a small area of the screen in the same direction. The majority of these sets of pixels (i.e., the linear motion

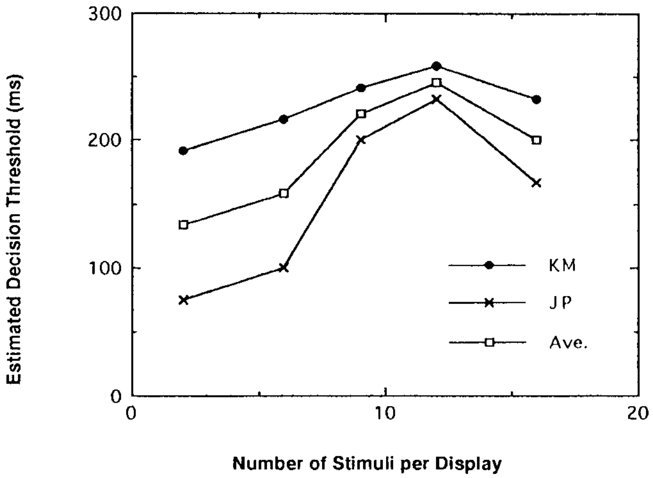

FIGURE 4. Results of a control experiment in which subjects identified the rotation direction of a rotating target. The 75% decision threshold is shown as a function of the number of stimuli per display. Data are shown for two subjects and their average.

distractors) showed identical motion, and the movements of the pixels contained in the rotating micro-displays differed from those of the distractors only in small variations in velocity and displacements on the *y*-axis. Under any circumstance, the rapid detection of the targets shown in this experiment is impressive; however, it cannot be guaranteed that such detection was not due to a difference in the two-dimensional appearance of the targets, relative to the general homogeneity exhibited by the distractors. Therefore, two subjects (the author and one of the paid assistants) replicated Experiment 1 with macro-displays that consisted of micro-displays containing different randomly positioned sets of pixels, thereby eliminating global homogeneity. The results of these two subjects were remarkably consistent with those they yielded in the main experiment: the percentage of correct responses never differed by more than 10% (i.e., 2 of 20 trials) for either subject in any condition. Therefore, it is suggested that the above results were not simply due to the perceptual appearance of 2D variations among the micro-displays. Nonetheless, a further experiment involving structure from motion shape discriminations was conducted and is reported below as Experiment 4.

In an effort to clarify the processing requirements of the task employed in Experiment 1, a second, related, task was examined. It was assumed that the detection of rotation (or linear translation) necessarily precedes its identification, and that identification of rotation direction requires a more effortful or advanced form of attention than detection. Therefore, it was thought that rotation direction identification would be sensitive to the number of stimuli present in a display. Two subjects (one previously tested, and one naive) were exposed to the rotating stimuli used in Experiment 1 under the same testing conditions described above. In this case, however, the subjects' task was to determine the direction of rotation (clockwise vs anticlockwise) of the target stimuli. Displays containing *only* linear movements, in the absence of a rotating stimulus, were not used. Each subject was tested on 30 trials in each condition. In order to maintain good accuracy along with speedy responses, subjects were given five trials in each condition with feedback prior to experimental testing (with two exceptions, condition accuracy in this session was maintained at 90–100%). As with the original data, 75% decision thresholds obtained during experimental trials (in which no feedback was provided) were subsequently determined. The results are shown in Fig. 4. At least two differences between the identification data shown in Fig. 4 and the rotation direction data of Fig. 3 are apparent. First, the average decision threshold shown in Fig. 3 was 127.73 msec, whereas for the data of Fig. 4 it was higher, 191.71 msec. Second, the data of Fig. 4 show consistent increases in decision threshold between displays containing two and twelve stimuli; the data of Fig. 3 tended to hover around the average decision time. Whereas the data of Fig. 3 could not be fitted well by any function other than a horizontal line, those of Fig. 4 show a significant linear trend between 2 and 12 stimuli: the slope of the best-fitting line was almost 12 msec/stimulus, and the line accounted for 95% of the variation in the data. As was the case in the original experiment, there was a decline in decision time for "filled" displays containing 16 stimuli. From the differences cited above, it seems reasonably safe to conclude that, as shown by the initial data of this experiment, the addition of linear motion stimuli to a display containing a single rotating stimulus does not add to the processing load required for detection of the rotating stimulus. Going beyond detection to a form of identification (i.e., naming rotation direction) appears to engage a more effortful form of attention. This general principle was also supported by the results of Experiment 4.

One comparison between the data shown in Figs 3 and 4 is consistent: in both cases, the estimated decision threshold declines between 12 and 16 stimuli per display. The reason for this decline is unclear, and discussion of the appearance of the displays with subjects did not show a subjective difference between displays containing 12 stimuli and those containing 16. A possible reason for this finding is that as all positions of the 4 × 4 grid become occupied with stimuli, the overall display approaches the display used by Dick *et al.* (1991) in appearance; that is, the irrelevant stimuli take on the character of a homogeneous background.

In both of the experiments reported thus far, when subjects were able to view the rotating stimulus centrally, or when a rotating stimulus in the periphery was viewed for a prolonged period of time, there was frequently a strong subjective impression of three-dimensionality, "objectness" and volume, along with rotation (this issues from spontaneous reports of the subjects as well as postexperiment questioning). Thus, to the extent that subjective reports are reliable, when display durations were long enough to permit the micro-genesis of a clear percept, observers appeared to have based their judgments on the gestalt of an object rotating in depth.

Nonetheless, it is possible that information secondary to the perception of rotation (e.g., deviations from purely linear translations of the pixels involved in rotation, pixel crowding at the edges of the rotating stimuli) provided reliable cues that the subjects might have used. In order to assess the extent to which such epiphenomena were used, two separate control experiments were conducted.

EXPERIMENT 2: TARGET LOCATION IN VISUAL FIELD

Although no track was kept of subject performance as a function of target location in the visual field in Experiment 1, on several occasions subjects noted that it seemed more difficult to be sure of judgments when the apparently rotating stimulus was in peripheral view. An examination of the displays showed that target stimuli appeared about equally in all locations in Experiment 1, so that the conclusions drawn above are not likely to be confounded by a procedural bias with respect to target location. Nonetheless, since the results of Experiment 1 reflect an averaging of performance across all target locations within the visual field, Experiment 2 was undertaken to determine the extent to which subjects could judge the presence or absence of a rotating target amidst a variable number of linear motion non-targets as a function of increasing distance from fixation. The number of stimuli per display was varied between two and five to determine whether small changes in this parameter would affect subject performance.

Subjects

Five subjects participated in this experiment. One subject was male (the author) and four were female subjects, two of whom were paid assistants and two of whom were unpaid volunteers. The female subjects ranged in age between 19 and 21 yr; the male was aged 41 yr. All subjects reported 20/20 vision and good depth perception, either with or without corrective lenses. When corrective lenses were indicated, they were worn throughout testing.

Stimuli and procedure

The micro-displays used in the present experiment were the same as those used in Experiment 1. However, the macro-displays differed in the following ways: first, four concentric circles were drawn around the fixation cross at the center of the display area. These had radii of 3.04, 5.47, 7.89 and 10.33 deg visual angle. In order to accommodate the large diameter circle and stimuli, the display area of the monitor was increased slightly. Second, in any macro-display, micro-displays were placed on one of the circles in random locations; their orientations remained the same as in Experiment 1. Third, the number of micro-displays contained in the macro-displays was either 1, 2, 3, 4 or 5. In all other respects, the procedure and details of Experiment 2 were the same as in Experiment 1.

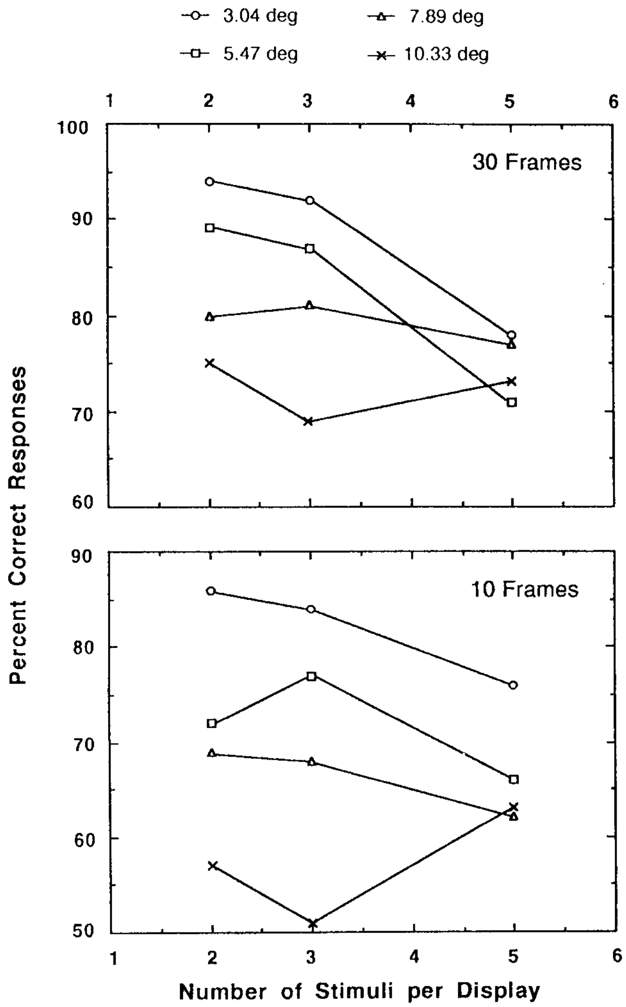

FIGURE 5. Results of Experiment 2, showing percentage of correct rotation detection judgments as a function of the number of stimuli per display. Distance of stimuli from fixation is the parameter. Separate graphs show the results from displays consisting of either 30 or 10 frames.

Results and discussion

As in Experiment 1, the percentage of correct decisions was not dependent on the presence/absence of a rotating stimulus. Therefore, the results are presented in Fig. 5, where the overall percentage of correct responses is shown as a function of the number of stimuli per display. Distance of the stimuli from fixation, in deg visual angle, is the parameter. Separate frames of the figure show the results obtained with displays that contained 10 and 30 frames. Apparent in Fig. 5 is the finding that accuracy was indeed influenced by distance from fixation. The data from this experiment were subjected to a 2 (number of frames) $\times$ 3 (number of stimuli) $\times$ 4 (distance from fixation) repeated measures analysis of variance. Distance of the stimuli was shown to have a significant main effect, $F(3,12) = 26.25$, $P < 0.001$. Similarly, as the relative elevations of corresponding curves in the two frames of Fig. 5 suggest, the number of frames per display also had a significant main effect, $F(1,4) = 15.43$, $P < 0.001$, with the 30 frame displays resulting in a higher percentage of correct responses than 10 frame displays. However, the number of stimuli per display did not have a significant main effect, $F(2,8) = 2.52$, $P > 0.05$, nor did it enter into a significant interaction with number of frames, $F(2,8) = 1.32$, $P > 0.05$, or with distance from fixation, $F(6,24) = 1.69$, $P > 0.05$. The three-way interaction was not significant, $F(6,24) = 0.24$, $P > 0.05$.

This pattern of results, along with those of Experiment 1, suggests the following interpretation. Both experiments show that when the percentage of correct responses is the dependent variable, increases in the number of frames per display result in increases in correct responding. However, the decision threshold data of Experiment 1 show that the stimulus duration required to reach 75% accuracy does not fluctuate systematically across conditions. This, along with the failure to find a significant interaction between number of frames and number of stimuli in both experiments suggests that beyond some minimum, increases in the number of frames serve mainly to provide opportunities for subjects to check initial impressions and/or refine their judgments, thereby improving the overall number of correct responses. Similarly, the failure of distance from fixation to interact with number of stimuli in the present experiment suggests that although the resolution of the rotation detection perceptual apparatus decreases with distance from fixation, the influence of competing stimuli does not make detection less efficient. The overall picture obtained from the conclusions of these two studies is that processing load (or attention required for detection) does not change with the number of linear motion stimuli contained in a display. Given that detection ability rarely approached 100% in the preceding two experiments, along with the fact that subjects did not find this to be an effortless task, it seems unwise to characterize the detection of rotation amidst linearly moving stimuli as involving a "pop-out" phenomenon. It is likely more consistent with both data and subjective experience to characterize the performance of this task as requiring a relatively constant degree of attention, regardless of the number of stimuli present.

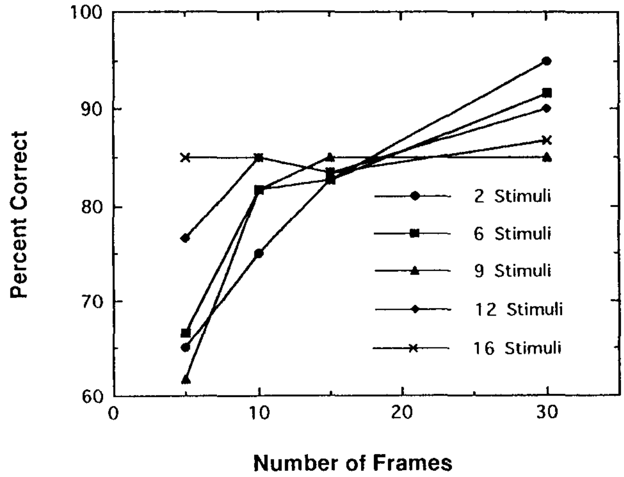

FIGURE 6. Results of Experiment 3, a replication of Experiment 1 with the exception that two columns of pixels were removed from the outer edges of the stimuli. Results show the percentage of correct rotation detection judgments as a function of the number of frames per display. Number of stimuli per display is the parameter.

EXPERIMENT 3: REMOVAL OF MICRO-DISPLAY EDGES

The present experiment was conducted in order to rule out the possibility that one, two-dimensional artifactual cue had been used previously to discriminate rotating from linear motion stimuli. Since the polar projection of a rotating cylinder produces crowding of pixels at its edges due to foreshortening, it is possible that the subjects in experiments 1 and 2 could have used the greater local density of pixels at the edges of the rotating cylinders to make their judgments, rather than the perception of rotation itself. To control for this possibility, 10 volunteer subjects participated in a preliminary experiment in which vertical edges of various widths (in pixels) were removed from both the rotating stimuli and linear motion micro-displays. Corresponding frames from pairs of rotating and linear motion micro-displays were then viewed for 3 sec each, with a 4 deg distance separating them (subjects stared at a fixation cross). On each trial, subjects rated the perceived similarity of the frames on a scale that ranged from 1 (identical) to 7 (extremely different). Five pairs of such stimuli, in which either 0, 1, 2, 3, 4, or 5 columns of pixels were removed from the outer (i.e., vertical) edges of the original micro-displays, were judged five times each by each subject. Average judgments applied to pairs' frames were compared to the expected value under the null hypothesis (i.e., 1) by single means t-tests. The results showed that the pairs of frames with 0 or 1 pixel columns removed were not perceived as significantly different; all others were. Therefore, in order to use stimuli in the present experiment that were as close as possible in detail to those used in Experiments 1 and 2 but which were nonetheless indiscriminable on a frame-by-frame basis, micro-displays with the outermost two columns of pixels removed from the originals were chosen.

The present experiment was a replication of the main Experiment 1, except that the micro-displays used to make the stimuli were modified as described above so as to make them indiscriminable when static. The goal of the experiment was to determine whether subjects still would be able to detect the rotating stimuli as well as in Experiment 1.

Subjects

Subjects were three of the participants who had originally served in Experiment 2.

Stimuli and procedure

Stimuli, apparatus, and viewing conditions were identical to those noted for Experiment 1, except for the fact that the micro-displays used to prepare the macro-displays now had two columns of pixels removed

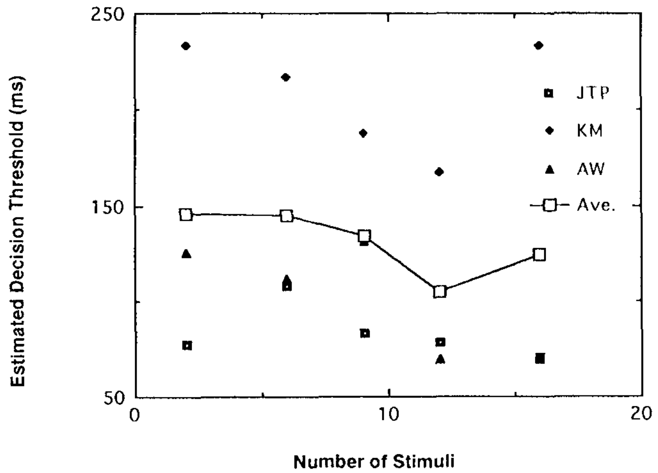

FIGURE 7. Results of Experiment 3, expressed as the 75% decision threshold as a function of the number of stimuli per display. Data are shown for individual subjects, along with their group average.

from the left and right edges of the cylinders and corresponding linear motion stimuli. The perceptual effect was the same as viewing the original micro-displays through an opaque mask hiding the outside edges. The procedure of the experiment was the same as in Experiment 1, except that the subjects viewed stimuli in each condition 40 times instead of 20.

Results and discussion

The percentage of correct responses as a function of the number of stimuli per display is shown in Fig. 6; number of frames per display is the parameter. As can be seen, there was a somewhat lower overall percentage of correct responses in the present experiment compared to Experiment 1. At the same time, the pattern of responses is similar between the two experiments, except that in the present experiment the curve for 16 stimuli remains flat. The results were subjected to a 4 (number of stimuli per display) × 5 (number of frames per display) repeated measures analysis of variance. This revealed a significant main effect for the number of frames, $F(3,6) = 6.61$, $P = 0.025$, but not for the number of stimuli, $F(4,8) = 1.99$, $P > 0.05$. The interaction between the two effects was also non-significant, $F(12,24) = 1.14$, $P > 0.05$.

As in Experiment 1, the 75% correct decision threshold was determined for each subject, and is shown plotted as a function of number of stimuli in Fig. 7. The average decision times obtained in the present experiment were very similar to those obtained in Experiment 1; that is, largely between 100 and 150 msec. As in Experiment 1, the function showing mean decision time showed no significant linear trend, $y = 151.41$ msec − (2.30 msec * x), $R^2 = 0.54$. On the basis of this experiment in comparison to the results of Experiment 1, it is, therefore, concluded that the crowding caused by foreshortening in the original rotation micro-displays contributed little, if anything, to the discrimination of those displays from others containing only linearly moving pixels. If anything, the crowding contributed to the development of a main effect of number of stimuli in the percentage correct data of Experiment 1. However, number of stimuli failed to affect the 75% decision thresholds of either Experiment 1 or the present experiment. Both experiments suggest that the degree of attention or processing required to detect rotating stimuli remains constant over the number of stimuli contained in a macro-display.

EXPERIMENT 4: SHAPE DISCRIMINATION IN STRUCTURE FROM MOTION

Do the results of Experiments 1–3 reflect the sensitivity of the human visual system to motion in depth and 3D structure, or are they simply a reflection of the ability of the SRP to detect small differences in the 2D local motion of the pixels that constitute target and distractor micro-displays? It was reasoned that if structure from motion based on the SRP is rapid and requires relatively little effortful attention to detect, then subjects ought to be able to detect particular rotating target shapes amidst rotating distractor shapes with about the same accuracy and speed that they detect rotating shapes amidst linear motion distractors. Whether the speed of such decisions varies with the number of distractors present ought to depend on the attentional requirements for the shape discrimination task, after the shapes have in fact been detected.

Experiment 4 was a replication of Experiment 1, with the exception that the target stimuli were rotating spheres, whereas the distractors were rotating cylinders (all of which always rotated in the same direction and with the same velocity). Also, to ensure that performance was not merely based on local 2D variations in motion, each micro-display had a different set of randomly positioned pixels.

Subjects

Subjects consisted of the author, one paid female assistant (who had previously served in Experiments 1–3), one unpaid male volunteer (age 19 yr), and one unpaid female volunteer (age 20 yr).

Stimuli and procedure

Stimuli were prepared along the lines of those described in Experiment 1 with the following exceptions: (a) instead of linear motion micro-displays, distractors now consisted of the same types of rotating cylinders used in Experiment 1; (b) these rotating cylinders, however, were not created from a single parent, but rather each was constructed with a new set of randomly positioned pixels; (c) target stimuli now consisted of rotating spheres, each constructed with a new set of randomly positioned pixels. The diameter of the spheres was increased relative to the cylinders by two pixels. This effectively reduced the appearance of "missing corners" without leading to the appearance of a noticeably larger figure. In all other respects, the stimuli were the same as described in Experiment 1.

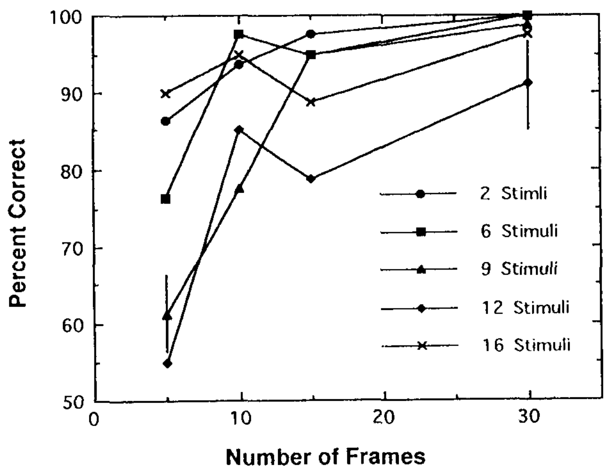

FIGURE 8. Results of Experiment 4, expressed as the percentage of correct rotation detection judgments as a function of the number of frames per display. Number of stimuli, or micro-displays, per display is the parameter. Error bars showed representative ±1 SE.

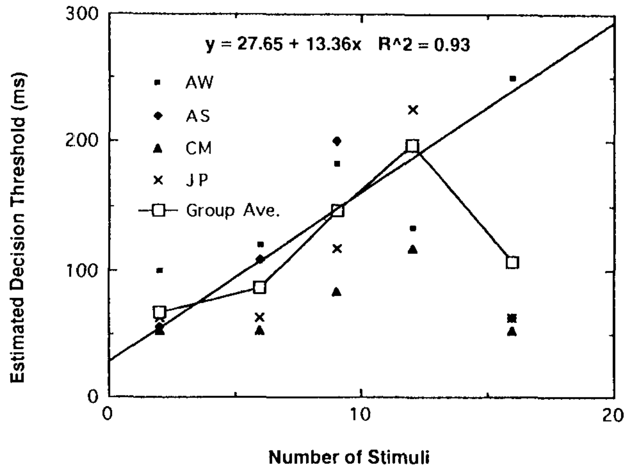

FIGURE 9. Results of Experiment 4, expressed as the 75% decision threshold (msec) as a function of the number of stimuli per display. Data are shown for individual subjects, along with the group average. Regression is shown for the first four average data points only. Note that the decision threshold for subject AW in the condition containing 12 stimuli is beyond the limits of the graph.

Subjects participated in a procedure that was identical to that described for Experiment 1, except now their task was to determine whether a macro-display did or did not contain a rotating sphere target.

Results and discussion

The overall percentages of correct responses as a function of the number of frames per display are shown in Fig. 8; the number of stimuli in each frame of the display is the parameter. Representative standard errors are also shown in Fig. 8. The smallest SE (0%) occurred in the following conditions: two stimuli/30 frames, six stimuli/30 frames, and six stimuli/15 frames. The largest SE (10.1%) occurred in the condition consisting of six stimuli/five frames. As was the case in Experiment 1, the more frames contained in a display, the greater was the percentage of correct responses (three exceptions occurred in the transition between 10 and 15 frame displays). The effect of the number of frames was again significant in a number of frames × number of stimuli repeated measures analysis of variance, $F(3,9) = 26.64$, $P < 0.001$. As in Experiment 1, the relationship between the number of stimuli contained in a display and the percentage of correct responses was non-monotonic: subjects were more accurate when displays contained 16 stimuli than they were when displays contained 12 stimuli, and in the 16 stimuli condition they were also more accurate with the 5 and 10 frame displays than in the 9 stimuli condition. The effect of the number of stimuli per display was also significant, $F(4,12) = 20.12$, $P < 0.001$. The interaction between the number of stimuli and number of frames conditions was also significant, $F(12,36) = 3.07$, $P < 0.01$. Post-hoc analysis suggests that the significant interaction was due to a failure of the 2 and 16 stimuli conditions to be influenced by the number of frames constituting the displays.

The percentage of correct responses per condition in Experiment 4 ranged from about 12% lower to about 5% higher than in Experiment 1, with the average percentage of correct responses being somewhat higher. Thus, given that subjects in the present experiment needed to rely on structure from motion information to make their judgments, it seems reasonable to conclude that subjects in Experiment 1 at least had structural information available with which to make their judgments.

As in Experiment 1, the next step was to estimate the 75% decision threshold for each subject. Of the 20 functions relating percentage of correct responses to the number of frames contained per display, two were non-monotonic and eight never fell below 75%. In these cases, estimates were made as described in Experiment 1. Figure 9 shows the resulting estimates of decision threshold as a function of the number of stimuli per display. The first finding of this experiment is that the decision thresholds are in the same general range (about 100–200 msec) as were the thresholds obtained in Experiment 1. This finding again suggests that the decisions made in Experiment 1 could have been based on structure from motion information. Additionally, Fig. 9 shows that there is a generally linear increase in decision threshold with increasing numbers of micro-displays for displays containing 2, 6, 9 and 12 stimuli (the regression shown in Fig. 9 is based on the first four conditions only). Once again a very low decision threshold was obtained when the screen was filled with 16 stimuli. Considering that the decision threshold obtained with the 16 stimuli displays may be anomalous, perhaps owing to the perceptual influence of a filled screen, these results suggest that the detection and discrimination of a target shape amidst similar distractor shapes (when both are defined by motion) may require a more effortful form of attention than the detection of a 3D rotation amidst linearly moving 2D stimuli.

Why should stimuli whose detection on the basis of rotation is relatively rapid, requiring little attentional control, be subject to an influence of the number of distractors when discriminations are made on the basis of shape? First, it has been known for some time that even when stimuli are defined by simple primitives (e.g., letters such as E, A and H which are defined by line segments), discrimination requires a longer reaction time the more similar the to-be-discriminated items become (e.g., E and F; Neisser, 1967). Thus, the cylinders and spheres used in Experiment 4, which were designed to be maximally similar, may have required a more effortful attentive processing to be discriminated because of structural similarity alone. Additionally, however, given that the cylinders and spheres used in this experiment were difficult to discriminate at all when frames were viewed statically, it is likely that the cylindrical and spherical forms themselves arose because of the activity of a "structure from rotation" process (suggesting that the detection of rotation in some sense precedes the formation of perceptual shape for these stimuli). It is possible that an "intra-channel" discrimination of shapes (i.e., discrimination of shapes defined by the same type of motion) requires a more effortful form of attention than an "inter-channel" discrimination (e.g., rotation vs linear motion).

GENERAL DISCUSSION

Insofar as possible, the stimuli used in the present experiments were designed to maximize the similarity between targets and non-targets. Stimuli were small areas that showed unidirectional motion of approximately 11 pixels. In Experiments 1–3, targets differed from non-targets only in the addition of path deviations that accommodated the "cosine factor" responsible for providing rotation information and the "perspective factor" responsible for providing rotation direction information (Braunstein, 1976). The only obvious non-motion-related cue that might have aided discrimination, edge crowding, was eliminated in Experiment 3 with little effect on the pattern of results or conclusions. Possible cues resulting from the identical positions of pixels in all micro-displays were eliminated in a small-scale control study associated with Experiment 1 and in Experiment 4. Furthermore, the motion of all of the pixels was nearly always short range. Under these conditions, when analyses were based on percentage of correct discriminations and the stimuli consisted of rotation amidst linearly moving distractors, the number of stimuli contained in a display produced a significant main effect only in Experiment 1, and that effect accounted for no more than a 15% absolute variation in correct discriminations (see Fig. 2). When analyses were based on the 75% decision time, the number of stimuli per display had no significant effects (except in Experiment 4, where the nature of the perceptual task differed), nor did it enter into any interactions. These results provide evidence that the process responsible for producing structure from motion operates very efficiently, perhaps pre-attentively, despite the extent of non-target linear motion. Additionally, when 3D rotation of a target shape needs to be detected and discriminated amidst other rotating 3D shapes, rapid detection appears to be modulated by a slower discrimination process (Experiment 4).

Although when the percentage of correct responses was the basis for analysis, the number of frames contained in a display had a consistent main effect, when this temporal factor was reduced to the time required to reach a 75% level of responding, the measure never varied significantly in any experiment where the subject's task was to detect rotation amidst 2D moving stimuli. This suggests that some minimum amount of processing time is required to recover structure from motion and that further viewing serves mainly to test the result, a conclusion that is consistent with the results of Liter *et al.* (1993) and of Treue *et al.* (1991).

The results of Experiment 4, in which subjects used structure from motion information to detect and discriminate target shapes (spheres) from distractors (cylinders) showed that subjects could perform the task accurately and with decision times that were comparable to those obtained in the previous experiments. This in turn suggests that 3D rotation and shape information was available to guide the decisions made by subjects in Experiments 1–3. However, the results of Experiment 4 also suggest that the time needed to make shape from motion judgments involving the detection/discrimination of a target shape amidst distractor shapes increases with the number of stimuli contained in a display, at least up to the point at which the screen becomes filled. Considering the similarity in appearance of the rotating spheres and cylinders that were used, this finding was not surprising: after (or perhaps concurrent with) the determination that 3D rotation was present in the displays, the task amounted to a fine-grained discrimination of the shapes of candidate objects.

Considered in total, the present results imply that when seeking a rotating sphere amidst linearly moving distractors, the detection of three-dimensionality alone is sufficient to guide responses. This process appears to be very rapid and to require a relatively low-level form of attention (i.e., one not seriously influenced by the number of distractors). When seeking a rotating sphere amidst rotating cylinders, however, structure from motion must give rise to percepts of at least two different shapes. This process occurs within the same broad time frame as the detection process itself, but decisions are influenced by the number of distractors present, indicating a possibly higher degree of attention investment.

The present results also address the question of the nature of the process responsible for the recovery of structure from motion. Specifically, the present results and conclusions are consistent with the interpretation of Dick *et al.* (1991), that under conditions of short-range motion, recovery of structure from 3D rotation simulations is a relatively low-level process that engages attention to approximately the same degree, regardless of the number of competing non-target stimuli. While not ruling out a role for the LRP, the work of Dick *et al.*

(1991), Mather (1989), Petersik (1991a), and Todd *et al.* (1988) suggests that the SRP makes a strong contribution to the recovery of structure from motion. These findings, together with an accumulated body of evidence that concludes that the SRP is itself a low-level, high-capacity, process (e.g., Dick *et al.*, 1987; Petersik, 1989), allow for the advancement of the hypothesis that the efficiency of rotation detection in the present experiment is due to the contribution of the SRP.

If it is true that the detectability of rotation amidst linearly moving non-targets is due to the activity of the SRP, then its efficiency may be in part due to a globally co-operative parallel-distributed type of processing. Previous results from this laboratory (Petersik, 1990) have demonstrated that the SRP behaves like a globally co-operative perceptual process when a collection of random dots is rotated about its center in the picture plane. If the same global co-operativity applies to the computations underlying the detection of rotation, it might explain the relative ease with which rotating stimuli are detected amidst competing noise stimuli. In fact, Treue *et al.* (1991) propose that structure from motion is the result of a global perceptual construction of a surface representation from global velocity measurements of moving elements, presumably reflecting the output of the SRP. Given the earlier research establishing that the SRP is co-operative in the processing of 2D displays, it is reasonable to hypothesize that the process responsible for the recovery of structure from motion in our rotation simulations is co-operative as well.

REFERENCES

Braunstein, M. (1976). *Depth perception through motion*. New York: Academic Press.

Dick, M., Ullman, S. & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, *237*, 400–402.

Dick, M., Ullman, S. & Sagi, D. (1991). Short- and long-range processes in structure-from-motion. *Vision Research*, *31*, 2025–2028.

Julesz, B. (1990). Early vision is bottom-up, except for focal attention. *Cold Spring Harbor symposium on quantitative biology* (Vol. LV, pp. 973–978). Long Island, NY.: Cold Spring Harbor Laboratory Press.

Liter, J. C., Braunstein, M. L. & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, *22*, 1441–1465.

Mather, G. (1989). Early motion processes and the kinetic depth effect. *Quarterly Journal of Experimental Psychology*, *41A*(1), 183–198.

Neisser, U. (1963). Decision-time without reaction-time: Experiments in visual scanning. *American Journal of Psychology*, *76*, 376–385.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703–1721.

Petersik, J. T. (1989). The two-process distinction in apparent motion. *Psychological Bulletin*, *106*, 107–127.

Petersik, J. T. (1990). Global cooperativity of the short-range process. *Perception & Psychophysics*, *47*, 360–368.

Petersik, J. T. (1991a). Effects of adaptation to apparent movement on recovery of structure from motion. *Spatial Vision*, *5*, 279–289.

Petersik, J. T. (1991b). Perception of three-dimensional angular rotation. *Perception & Psychophysics*, *50*, 465–474.

Petersik, J. T., Pufahl, R. & Krasnoff, E. (1983). Failure to find an absolute retinal limit of a putative short-range process in apparent motion. *Vision Research*, *23*, 1663–1670.

Petersik, J. T., Shepard, A. & Malsch, R. (1984). A three-dimensional motion aftereffect produced by prolonged adaptation to a rotation simulation. *Perception*, *13*, 489–497.

Shulman, G. (1991). Attentional modulation of mechanisms that analyze rotation in depth. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 726–737.

Todd, J. T., Akerstrom, R. A., Reichel, F. D. & Hayes, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics*, *43*, 179–188.

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

Treue, S., Husain, M. & Anderson, R. A. (1991). Human perception of structure from motion. *Vision Research*, *31*, 59–75.

Verghese, P. & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, *34*, 2453–2467.