

The Perception of Structure from Visual Motion in Monkey and Man

R. M. Siegel¹ and R. A. Andersen²

Developmental Neurobiology Laboratory
The Salk Institute for Biological Studies

Abstract

■ The ability to perceive structure using motion information was examined using a reaction time task with two primate species. *Homo sapiens* and *Macaca mulatta* subjects were quantitatively tested under identical conditions to detect the change from a control unstructured to a test structured motion stimulus. The structures underlying the test were rotations of a plane, expansion of a plane, and a rotation of a three-dimensional cylinder. On many of the stimulus conditions, the two

species performed similarly, although there were some species differences. These differences may be due to the extensive training of the monkeys or the use of different cognitive strategies by the human subjects. These data provide support for the existence of a neural mechanism that uses flow fields to construct two- or three-dimensional surface representations. ■

INTRODUCTION

The correct analysis of motion by the primate visual system is ethologically important. Nonhuman primates apparently can use motion cues well enough to guide themselves through the jungle environments at a fast pace (Bertrand, 1969). The importance of motion in man for determining the shape of objects and discrimination of items in complex visual scenes has been described (von Helmholtz, 1910; Beverly & Regan, 1975; Wertheim, Wagenaar, & Leibowitz, 1982). The analysis of visual movement for both species can be separated into "low-level" and "high-level" processing. Low-level motion concerns the initial extraction of the direction and the speed of moving points in the visual field (Braddick, 1980). High-level motion is more concerned with extracting spatial features of the environment and different cues have been categorized. For example, on optical and physical grounds Longuet-Higgins and Prazdny (1980) noted that motion cues consist of (1) expansions and contractions of points in the visual world around a central point, (2) rotations of the world about a central point, and (3) translations. Other motion cues have been defined (e.g., shearing motion: Nakayama & Tyler, 1981).

Motion information can be used for either object identification (Wallach & O'Connell, 1953; Johanssen, 1982) or location of the body in the environment (Longuet-Higgins & Prazdny, 1980); Wallach and O'Connell (1953) were able to show psychophysically that human subjects could correctly determine the three-dimensional shape of rotating visual objects using monocularly presented

visual motion. They termed the illusion the "kinetic depth effect." More recently the general problem of determining three-dimensional shape from motion has been termed the "structure from motion problem" (Ullman, 1979).

An understanding of the cortical areas and pathways that are involved in the analysis of visual motion has proceeded in parallel with the above work through studies in the anesthetized monkey. Direction of movement selective cells with approximately 1° diameter receptive fields in macaque visual cortex were first described by Hubel and Wiesel (1977). It was later shown that several extrastriate visual areas also contain neurons that are selective for visual motion; the most studied of these areas is the middle temporal area (MT or V5). Area MT has a retinotopically organized representation of the contralateral visual field in owl monkey (Allman, Kaas, & Lane, 1973; Zeki, 1980, Petersen, Baker, & Allman, 1980) and macaque monkey (Zeki, 1974; Van Essen, Maunsell, & Bixby, 1981; Maunsell & Van Essen, 1983a). The sizes of the receptive fields of MT neurons are larger than those of striate cortex and vary in proportion to eccentricity. Furthermore this area has been shown to have a columnar organization for direction of movement (Albright, Desimone, & Gross, 1984). Area MT projects to area DP and the medial superior temporal area (MST or V5a) in the owl monkey and macaque monkey, respectively (Baker, Petersen, Newsome, & Allman, 1981; Maunsell & Van Essen, 1983b; Ungerleider & Desimone, 1986a, b). In contrast to area MT, the receptive fields in these putatively homologous two areas are larger and can be

bilateral (Mikami, Newsome, & Wurtz, 1986; Desimone & Ungerleider, 1986). In the anesthetized macaque, cells have been found in MST that are sensitive to planar rotation, expansion, and three-dimensional rotation of objects (Saito, Yukie, Tanaka, Hikosaka, Fukada, & Iwai, 1986; Tanaka, Hikosaka, Saito, Yukie, Fukada, & Iwai, 1986). There are other areas that may also process motion (e.g., 7a: Motter & Mountcastle, 1981; Sakata, Shubutani, Ito, & Tsurugai, 1984; Sakata, Shubutani, Kawano, & Harrington, 1985). How these different motion cells acquire their different properties and how the properties contribute to motion perception are still unanswered questions.

There has also been a strong interest in computational and machine vision in determining algorithms that can correctly adduce the structure of the environment from motion (Ullman, 1979; Bennett & Hoffman, 1985; Zucker, 1983, 1985; Hildreth, 1983, 1984; Grzywacz & Hildreth, 1987; Grzywacz, Hildreth, Inada, & Adelson, 1988). This work has attempted a "top-down" view by describing the "computational" or "functional" aspects of the problem (Marr, 1982). There have been essentially two approaches to this question. First, algebraic proofs have been constructed that demonstrate the minimal conditions needed to extract the location of points in the environment from motion input (Ullman, 1979, Bennett & Hoffman, 1985; Longuet-Higgins & Prazdny, 1980; Cornilleau-Peres & Droulez, 1987). Ullman's structure from motion theorem for rigid objects is an example of this approach. Second, algorithms have been described that iteratively attempt to model the correct structure of the environment and match this structure to the actual input motion flow field (Hildreth, 1983, 1984; Zucker 1983, 1985; Grzywacz & Hildreth, 1987; Grzywacz et al., 1988). Implicit in both of these approaches are attempts to implement known information (or constraints) about the physical and optical environment. Two important results from these studies are that the number of points identifying the object (Ullman, 1979) and the amount of time the points are displayed (Zucker, 1983, 1985) are important in extracting structure from motion.

We have been conducting experiments on human and nonhuman primates that concern their ability to perform "high level" motion analysis. The ultimate aim of this work has been to combine the psychophysical, physiological, and computational approaches in studying visual motion processing. With this in mind, we have designed new motion stimuli that are highly controlled to study high-level motion analysis. We have restricted ourselves to examining three types of movement: planar rotations, expansion/compression, and three-dimensional rotations. In the present work, the stimuli are described and the ability of both species of primates to perform these motion analyses is characterized in both the temporal and spatial domains. Briefer reports of parts of this work have been presented previously (Andersen & Siegel, 1986; Andersen & Siegel, 1990; Siegel & Andersen, 1988).

RESULTS

The structure from motion reaction time task used in these experiments consisted of the subject moving a lever in response to a display of 128 moving points. The beginning of each display consisted of a control unstructured motion stimulus (see Materials and Methods) that changed at a random time to a structured motion stimulus. The trajectories in the structured motion stimulus were determined by the movements corresponding to (1) a plane *rotating* about a normal to its surface, (2) a surface *expanding*, or (3) the *three-dimensional rotation* of a cylinder (Fig. 1). The control unstructured stimulus was generated by spatially mixing the appropriate motion trajectories (Fig. 2, see Material and Methods). The subject pulled back a lever at the onset of the unstructured stimulus and released it after the change to structured motion. These stimuli were highly controlled so that the only cue for the subject to release the lever was this change from unstructured to structured motion.

Initially monkey subjects were trained in a dimming task (Motter & Mountcastle, 1981). Then they were trained to perform the structure from motion tasks. The monkey subjects were then trained over a period of days to weeks (approximately 10,000 to 50,000 trials for two monkeys, only 300 trials for M-60). Human subjects, in contrast, had a number of chances to practice the task, but none viewed the stimuli as many times as the monkey (approximately 100 trials of training). The results using these three different types of stimuli are described separately. In the studies to be described below, (1) 128 points are used in each display, (2) the point life is 532 msec, (3) the display rate is 35 Hz (interframe interval 24 msec), and (4) the display is 15° in diameter unless otherwise stated. The effect of simpler motions (i.e., rotation and expansion/compression) is considered separately from the more complex three-dimensional

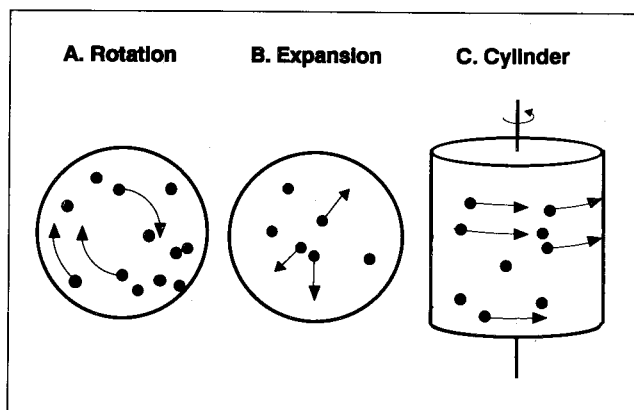


Figure 1. Diagrams of the three motion fields used. (A) Rotation in a plane. Each point in the motion flow field rotates about a central point. (B) Expansion in a plane. Each point moves radially away from the center point. In this display, the radial velocity is independent of the radial distance. (C) Rotation in three dimensions (the hollow cylinder). A cylinder is rotated about a vertical axis.

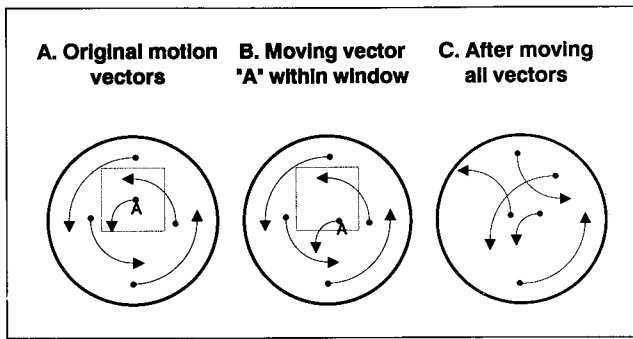


Figure 2. Generation of the unstructured motion displays. The movement trajectory of each point is indicated by the arrows. (A) The stimuli are generated by first computing the location of points located on the surface of the rotating plane for each instant in time. This display is called the *test display*. Note the box drawn about point A. (B) To unstructure the motion display, the entire trajectory associated with point A is randomly displaced within the region marked by the box. (C) The *control display* is computed by repeating this process for all of the points' motion trajectories from the test display. The fraction of structure associated with the display is given as a function of the ratio of the box and the diameter of the display. The point density on the surface of the display plane is kept constant; each point is displayed for a finite amount of time. At the end of this time, the point vanishes and reappears randomly on the screen to follow a new motion trajectory.

structure from motion due to a clear-cut division in the subjects' performance with these stimuli. As well there is a conceptual division between these two sets of stimuli. The visual system of the subject is extracting a three-dimensional shape only when the stimulus is a cylinder.

Rotation, Expansion, and Compression Structures from Motion

Planar Rotation

The subjects were first tested on a detection of the change from unstructured to structured rotation in a plane. All monkey and human subjects were able to detect this change with 97–100% of the trials correct. Psychometric functions were obtained for the subjects. Figure 3A shows such data for subject monkey 61. It can be seen that as the fraction of the structure of the display is decreased, the subject gets a lower percent of the trials correct. The mean and asymptotic standard for the fraction of structure required for the subject to get 50% of the trials were computed for each individual and is shown in Figure 4A.

Similar curves were obtained for the other human and monkey subjects. The thresholds for all the subjects tested appeared comparable (Fig. 4). Statistically the 50% thresholds were compared using a two way analysis of variance (Table 1) with groupings by species and stimulus type (see below for description of expansion data). Neither stimulus nor species had a significant effect on the 50% threshold indicating that both species performance was equivalent for both stimulus conditions. The

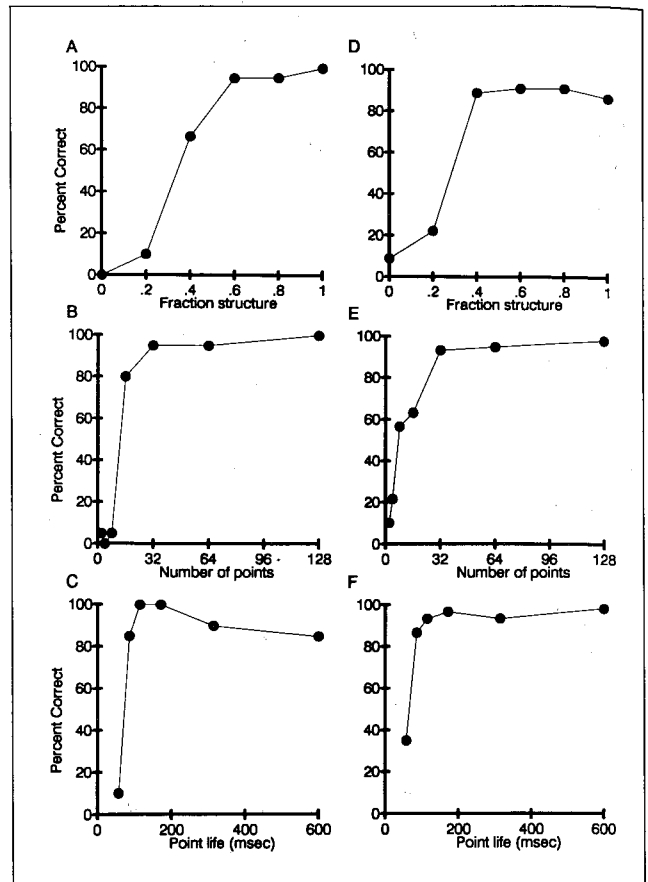


Figure 3. Performance of monkey 61 at detecting the change from unstructured to structured motion for rotation and expansion motions. (A–C) Obtained using planar rotation; (D–F) Obtained using expansion. (A, D) Varying fraction of structure. Note that for the lower displayed structure, the monkey's performance decreases. (B, E) Number of points. Decreasing the number of points made it more difficult for the animal to perform the task. (C, F) Decreasing the point life also made it more difficult to perform the task. The vertical axis is the percent of trials performed correctly.

mean and standard deviation of these 50% thresholds for the rotational stimulus were 0.233 ± 0.052 for the five subjects.

Both human and monkey subject's performance was dependent on the number of points and the point life (Figs. 3B and 4B). In these experiments, the fraction structure of the display was kept constant at 0.875 and one of the other parameters was varied. For the planar rotational display, all of the subjects' performance decreased with the point life when the number of points were kept constant at 128 (Figs. 3B and 4B). We computed the 50% thresholds using linear interpolation since the dependence of the percent correct on the varied stimulus parameter had a hump or a sharp bend unlike the logit function. Using this analysis, the 50% threshold for the rotational dependence on the point life was 68.7 ± 4 msec. On the other hand there were clear differences in the dependence of the monkey and human subjects on the number of points (Figs. 3C and 4C).

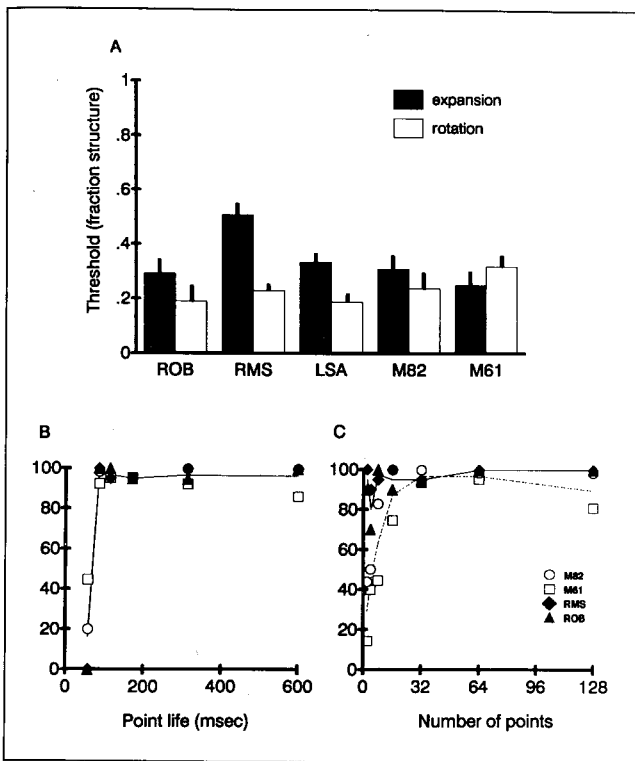


Figure 4. Thresholds for rotation and expansion structures from motion for all subjects. (A) The threshold needed to get 50% of the trials correct for detection of rotation (open bars) and expansion (filled bars) is shown. The differences were not significantly different using an analysis of variance (see text and Table 1). The error bars represent the asymptotic standard deviation (Ashton, 1972). (B) The effect of point life on the subjects' performance for the rotation stimulus. If the point life was less than 100 msec, it was difficult for all the subjects to perform the task. The solid line is the average value for all subjects. (C) The effect of number of points for the rotational stimuli. The human subjects required fewer points than the monkey subjects. The human subjects reported that they could look for the points moving either about, or not about the center, for low number of points. The solid line is the average for the human subjects; the dotted line is the average for the monkey subjects. (See text for statistical analysis.)

Human subjects were able to perform the rotation task at over 90% correct with as few as two points (Fig. 4C solid line) while the monkey subjects performance dipped below this level for 16 or less points. The monkey 50% threshold that was computed using the linear interpolation was 6 ± 5 points; it was not possible to compute a 50% threshold for the human subjects. The human subjects queried with respect to their strategy under the few point condition indicated that they were actively looking for the matching centers of rotation for each of the two points. It may be that the monkey subjects were unable to make this abstraction.

A χ^2 test for two independent samples (Siegel, 1956) was used to statistically compare the two species performances on the number of points and point life. The number of correct trials was computed for each subject under each condition and then averaged across species

members. The performance was statistically different for the number of point dependency [$\chi^2(6) = 16.6, p = 0.01$]. For the point life dependency, the null hypothesis that there were no differences by species could not be rejected ($\chi^2(5) = 4.34, 0.2 < p < 0.3$).

Expansion

Similar results were seen using the expansion stimuli (Figs. 3 and 4). All of the human and monkey subjects were able to detect the change from unstructured expansion to structured expansion at a level of 95–100% correct. A psychometric function, with the independent variable being fraction of structure, was collected for each subject. It was found that the thresholds again were similar (Fig. 4A), being on average 0.336 ± 0.102 . The dependence on number of points and point life was examined for some of the subjects. As with the rotation task, fewer points and shorter point life led to fewer correct trials as seen in Figure 3E and F. However an insufficient number of subjects were tested using these last two conditions to make any quantitative statements about the similarity of their performance.

These results indicate that both species are able to distinguish the difference between unstructured motion and structured motion. Furthermore both species ability to perform this task is similar except in the case of number of point sensitivity for the planar rotation stimulus. The impaired performance of the monkey subjects relative to the human subjects may be due to the human subjects ability to use more complex cognitive strategies to perform the task. Although the displays were highly controlled with respect to point density and form cues, it is possible to assume that the task could be performed using less obvious local *motion* cues (e.g., crossing of motion trajectories). Control experiments addressing this possibility will be described below.

Three-Dimensional Structure from Motion: Hollow Cylinder

In this study, the display was square and 7.5° on a side. This size display was used rather than the larger 15° display as preliminary experiments indicated some human subjects had difficulty with the larger display. In this study, the reaction time window was modified from its normal value of 150 to 800 msec to 300 msec to 1200 msec. This was necessary since we observed early that the subjects were consistently releasing the key late with the cylinder stimulus.

All subjects were able to detect the change from unstructured to structured motion when the underlying three-dimensional structure was a hollow cylinder. This is the first time to our knowledge that nonhuman primates have been shown to be able to use the motion flow field intrinsic to a moving three-dimensional object to perform a behavioral task. All monkey subjects were

Table 1. Two-Way Analysis of Variance for the 50% Threshold to Rotation and Expansion Stimuli^a

Source of error	df	Sum of squares	Mean squares	F	p
Species	1	3.57	3.57	0.06	0.811
Stimuli	1	179.08	179.08	3.13	0.127
Interaction	1	179.08	179.08	3.13	0.127
Error	6	342.93	57.15		
Total	9	794.21			

^aThe species were human and monkey subjects; the stimuli were planar rotation and expansion. Neither of the two factors was significant in determining the 50% threshold.

able to perform this task within 50–100 trials. Psychometric functions were determined for the dependence on fraction structure, number of points, and point life (Fig. 5).

It was found that the 50% threshold of fraction of structure (being 0.541 ± 0.161 for all five subjects tested) needed to perform the three-dimensional task was greater than for either of the other stimuli (0.233 ± 0.052 and 0.336 ± 0.102 for rotation and expansion, respec-

tively). This suggests that more complex processing of the visual input is needed to perform the three-dimensional task (cf. Figs. 3 and 5). The 50% thresholds for the three-dimensional stimuli were not significantly different for the two species when compared using an analysis of variance ($p = 0.266$, $n = 3,3$). The mean and standard deviations were 0.62 ± 0.04 and 0.46 ± 0.21 for the human and monkeys, respectively. The grand mean and standard deviation was $0.54 \pm .16$ ($n = 6$).

The increased difficulty of the task is supported by the subject's dependence on the number of points and point life (Fig. 5B and C). In the point life study, the points were presented at 70 frames per second rather than 35 frames per second to achieve greater temporal resolution. The subjects always needed more points and longer point lives to perform the three-dimensional task than the two-dimensional task.

Statistically the dependence on the point life and number of points was tested for species effects using the independent χ^2 test as described above. The human and monkey performance was not significantly different for the number of point dependence [$\chi^2(7) = 13.6$, $0.05 < 0.10$]; the species had different dependencies for the point life [$\chi^2(6) = 25.1$, $p < 0.001$].

CONTROLS

The above results suggest that both human and monkey subjects were able to detect the change from unstructured to structured motion when the underlying structure has either two or three dimensions. Human subjects reported a strong sense of seeing a surface or object being moved and indicated that the percept was obtained from the entire display. However, as indicated above, it is possible to perform the rotation and expansion tasks using only local motion cues. It is also possible in theory to perform the three-dimensional task using local cues in the following way. Early the subject looks for an inconsistency in the motion vectors by scrutinizing one small local region of the display and releases the key when the distributions of velocities in this region become smaller. It would seem unlikely that this cue is

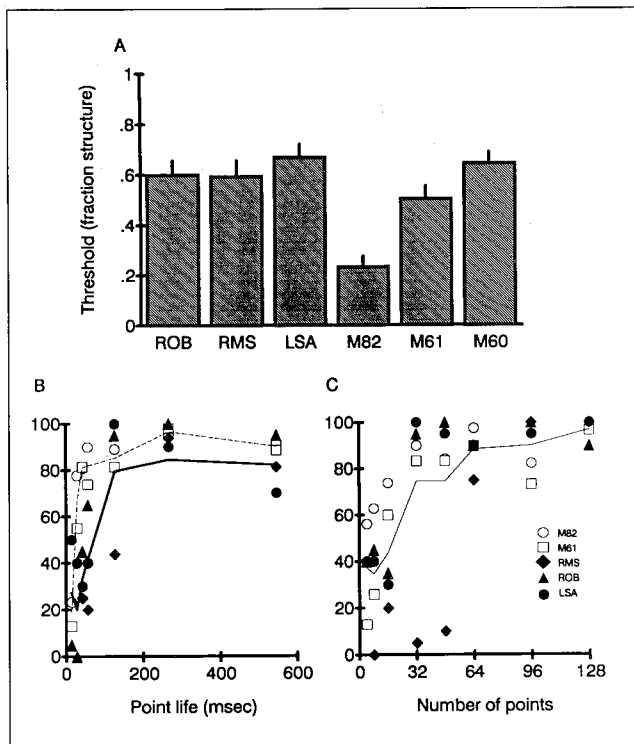


Figure 5. The detection of the change from unstructured to structured motion using the hollow cylinder display for monkey 61. (A) Thresholds (mean and asymptotic standard deviation) for all subjects for the cylinder stimulus obtained using logit analysis. Note that the threshold (0.541 ± 0.161) is greater than the simpler motion displays. (B) Point life. The subject can perform when the point life is at least 100 msec. (C) Number of points. More than 32 points are needed to perform this task.

useful to extract structure from motion as the number of points viewed locally are too few to perform the task from the above results on point number. None of the human subjects queried after the study reported using this strategy. However, we could not be so sure for the monkey subjects. Thus a control experiment was performed to test if local cues were being used by either species.

In this study, the display was masked so that only 1/25th of the display by area could be observed (3° square for the two-dimensional display, 1.5° for the three-dimensional display). On the average about five points are viewable at any one instant in time in the masked display. Psychometric functions were obtained for the subjects with the standard parameter set. It was found that all subjects were able to perform the two-dimensional task with the masked display. However, their thresholds were markedly increased (Fig. 6). For the expansion stimulus the 50% thresholds without and with the mask were 0.339 ± 0.099 and 0.609 ± 0.0588 , respectively. For the planar rotation stimuli, the 50% thresholds without and with the mask were 0.235 ± 0.054 and 0.405 ± 0.045 , respectively. This suggests that the two-dimensional structure can be detected using local cues, however global cues result in better performance. The most obvious local cues to use are local parallelisms.

All subjects were unable to detect the change from unstructured to structured three-dimensional motion when the display was masked. As psychometric functions were not obtainable under these conditions the subjects' performance with and without the mask is given in Figure 6. Without the mask in place, the average percent correct for the five subjects was $95.9 \pm 2.90\%$; with the mask in place the average percent correct was $18.6 \pm 21.1\%$. This result indicates that the ability to obtain three dimensional structure from motion requires a global analysis of the display unlike the two-dimensional display.

Reaction Times

Analysis of the time from stimulus change to key release further suggested some constraints on the process of abstracting structure from motion. The first point to note is that, for all subjects, the reaction time is about 200 to 600 msec greater for the structure from motion tasks than that found when doing a simpler motion detection task for a shear stimulus (cf. present work and Golomb, Andersen, Nakayama, MacLeod, & Wong, 1985). This result suggests that the structure from motion task uses cortical areas beyond that needed for simple motion detection (see Discussion). We next will consider differences found in abstracting rotation, expansion, and three-dimensional flow fields. The human and monkey data will be described separately as there are some species differences.

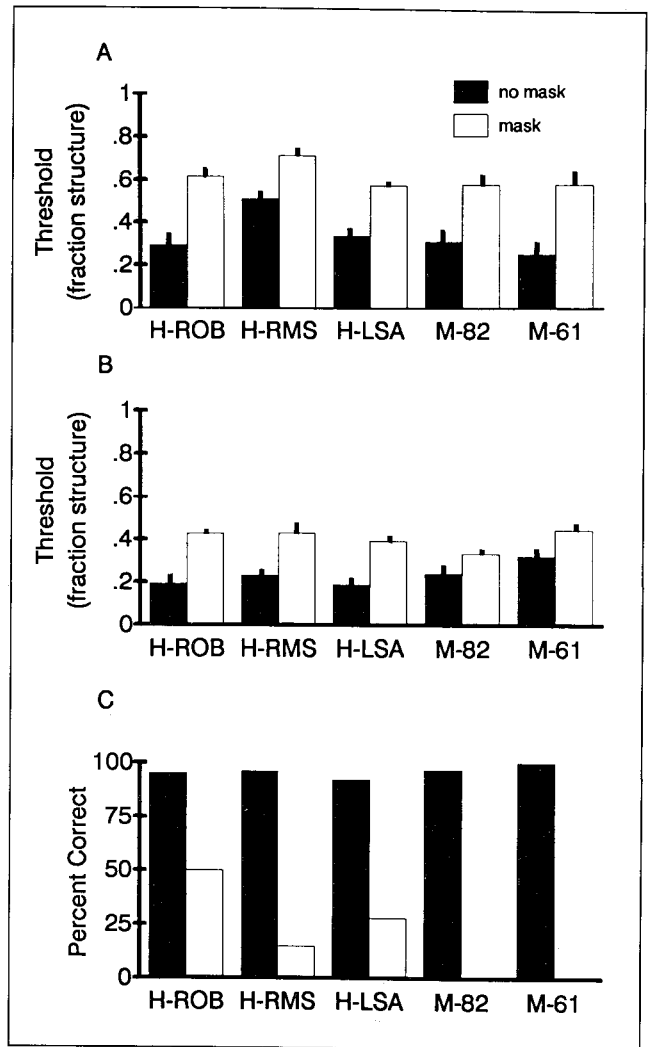


Figure 6. The effect of masking 1/25th of the display by area. The thresholds to planar rotation (A) and expansion (B) were increased. It was impossible to perform the task for the hollow cylinder display (C) so the thresholds could not be computed. Instead the subjects' performance for the completed structured display (fraction structure is 1) is indicated for the unmasked and masked displays. The open bars indicate the unmasked values while the closed bars indicate the masked values. Note subjects M-82 and M-61 got 0% trials correct when the display was masked.

Human Subjects

A summary of the human reaction times for these stimuli (the fraction of structure was 1; the number of points was 128; point life was 532 msec) is presented in Figure 7. It can be seen that for all three subjects, the reaction time for the cylinder task was greater than that with either the rotation or expansion task. A two-way analysis of variance was performed on the reaction time data using subject and stimulus as the dependent variable (Table 2). There was no interaction between these two groupings, however it can be seen that both subject and stimulus are main effects.

Figure 7. Reaction times of the human and monkey subjects when the display changed to complete structure. Each of the human subjects had a longer reaction time for the hollow cylinder display as opposed to the planar rotation and expansion display. One of the monkey subjects had longer reaction times for the hollow cylinder. Two of the monkeys had equal reaction times for all three stimuli.

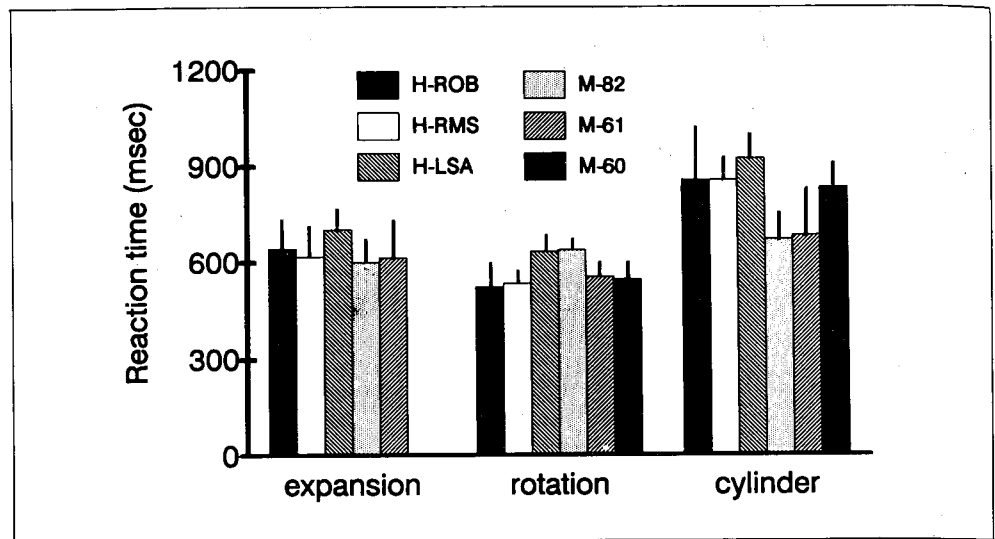


Table 2. Two-Way Analysis of Variance for Human Subjects^a

Source of error	df	Sum of squares	Mean squares	F	p
Subject	2	336800	168400	21.6	0
Stimuli	2	3899500	1949700	251.0	0
Interaction	4	22933	5733	0.738	0.867
Error	216	1677700	7767		
Total	224	5936900			

^aThe data were taken from the mean of 25 trials where each subject viewed a stimulus with fraction of structure of 1, number of points 128, and point life 532 msec. It can be seen that differences of reaction time exist both by subject and by stimulus (*p* values of 0 indicate *p* < 0.001).

A longer reaction time was needed to perform the cylinder task as compared to the rotation or expansion tasks (Fig. 7). For comparison purposes, the mean reaction times for the human subjects were computed as 656 ± 93 , 565 ± 76 , and 878 ± 114 msec for the rotation, expansion, and cylinder stimuli, respectively. The cause of the shorter time needed to perform the expansion task versus the planar rotation task remains to be explored in future work. It is also interesting to note that the technique used here is precise enough to find differences between human subjects.

There are a number of possible sources for this increased reaction time in the three-dimensional rotation versus the rotation and expansion stimuli. The first was that there were two surfaces (front and rear) delineated by motion in the three-dimensional display. As a control for the number of surfaces being the source of the longer reaction time, we tested the subjects with a planar rotation display where there were two surfaces—one moving clockwise and the other counterclockwise (Fig. 8). All subjects were able to detect the change from an unstruc-

tured to structured stimulus at close to 100% trials correct with this stimulus.

A two-way analysis of variance for these stimuli (not shown) demonstrated that there were both highly significant subject and stimuli differences as well as a significant interaction terms. The clockwise/counterclockwise planar rotation stimulus led to a reaction time that was about the same as the clockwise or counterclockwise rotation stimuli. Thus it appears that the longer reaction time found when viewing the cylinder stimulus was not a result of the stimulus consisting of two surfaces.

Another possible source for the longer reaction times with the three-dimensional stimuli is that they are only due to the time that it takes for the stimulus to change. However the time needed to change is identical for all stimuli. Furthermore the reaction time is rather independent of the point life over a large range. The reaction time remains long even as the point life is decreased (Fig. 9). Indeed the reaction time was not strongly dependent on the main three parameters of the display (i.e., fraction of structure, number of points, and point

Figure 8. Comparison of reaction time for a clockwise (right) planar rotation, a counterclockwise (left) planar rotation, simultaneous rotations, and the cylinder stimulus. The reaction time was no greater for the simultaneous rotation than for either the right or left rotation for all subjects.

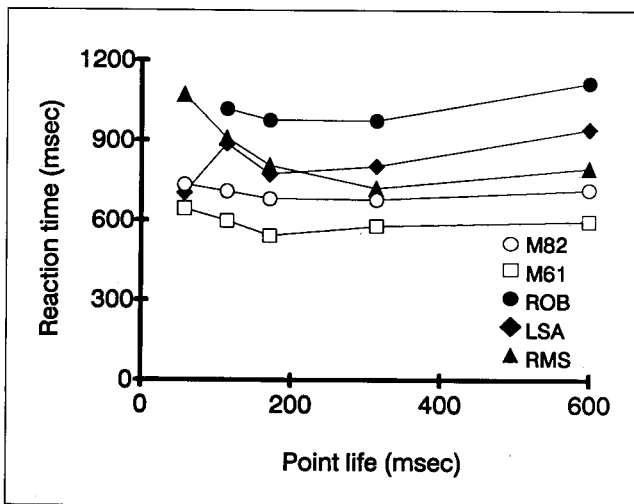
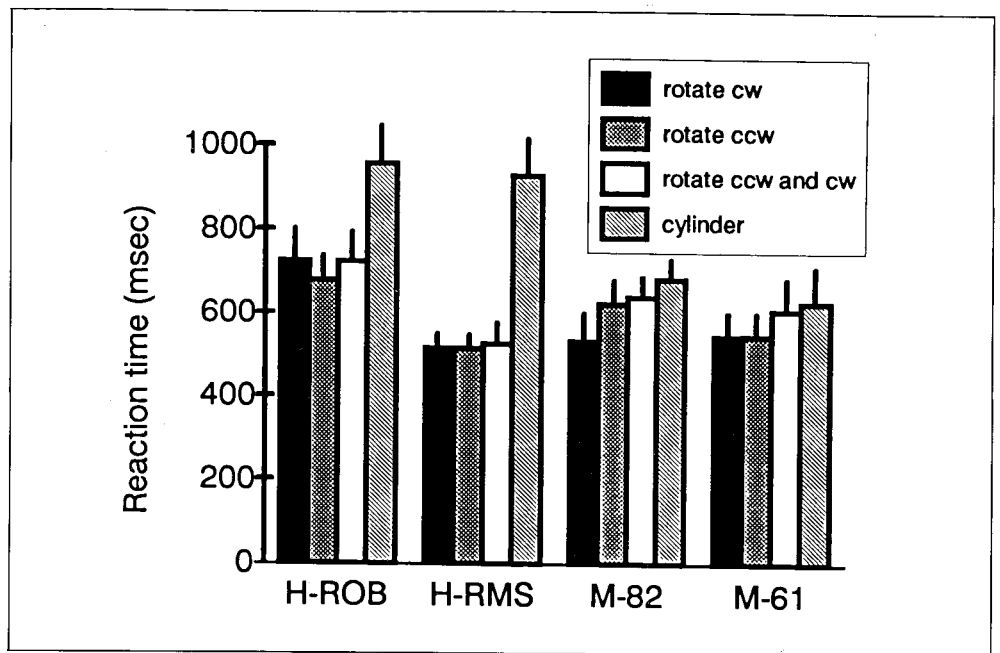


Figure 9. Reaction times of the human and monkey subjects as a function of the point life of the display. As the point life is increased there is little increase in the reaction time. Therefore the time for the display to change is not the main factor in determining the reaction time.

life). Thus the longer reaction time found with the cylinder stimulus is suggestive that additional neural processing is required in the human subjects to solve this task.

Monkey Subjects

One of the monkey subjects (M-60) had an obvious difference between his rotation and expansion reaction

time versus his three-dimensional reaction time in a manner similar to the human subjects (Fig. 7). However, the two other subjects (M-82 and M-61) did not have such a difference. The difference between monkey subjects may arise from training at the task as (1) the reaction time for the cylinder display dropped for M-82 and M-61 over the period of training and (2) M-60 viewed the display only 300 times. (It is interesting to note however that the most experienced human subjects were unable to reproduce these short reaction times.)

A two-way analysis of variance was performed for the M-82 and M-61 (not shown) with subject as one grouping and stimulus type as the other grouping variable. There was a significant interaction term ($p = 0.02$). Stimulus type had no effect ($p = 0.42$) while stimulus had a significant effect ($p < 0.001$). The reaction times were averaged across the two monkeys and were 608 ± 90 , 598 ± 64 , and 678 ± 117 msec ($n = 50$ trials) for the rotation, expansion, and cylinder tasks, respectively.

Like the human subjects, a relative invariance of the reaction time with the difficulty of the task was found with the monkey's performance (Fig. 9). This suggests that the long reaction time is not due only to the time needed for the change. It is also important to note that the reaction times for all the subjects were longer than that measured in simple dimming tasks or saccade tasks where the reaction times are on the order of 300 msec. This indicates that for all the motion stimuli described here, additional processing is required to perform these integrative tasks.

DISCUSSION

Comparison of Performance for the Two Species

The present results indicate that human and nonhuman (*Macaca mulatta*) primates have similar performances for rotation and expansion motion perception tests although there are notable species differences. The subjects' performances were similar for many of the parameters tested. Both species dependency on the masking of the display was the same. As well both species had similar reaction times for the simpler two-dimensional motion stimuli. However, it is crucial to note that some differences were found. Human subjects did quite well on the planar rotation task when only a few points were present as compared to the monkey subjects. This may reflect the use of a cognitive strategy too abstract for the lower primate to discover (or different from that used by the monkey.)

In view of the similarities between the two species, neurophysiological study of the computation of global motion by the visual system of the *Macaca mulatta* to understand the process in *Homo sapien* is justified. During such study, however, it will be necessary to be aware of the differences between the species and see if physiological explanations for them can be found. It is also clearly necessary to delimit the source of these differences (e.g., training effects) through further psychophysical studies.

A similar statement can be made with respect to the three-dimensional motion stimulus (cylinder) with the following provisos about the temporal aspects of processing. First, two of the three monkeys tested could do the task with shorter life times than the human subjects. Second, for the same two monkeys, the reaction time to detect the cylinder was shorter than any of the other subjects. As noted in the Results section, one possibility is additional training. However, no human subject to date, however experienced, has been able to duplicate this performance. Two other possibilities for this difference between monkey and human and subjects is (1) variation of performance among individuals or (2) there is indeed some difference in the ability of human and monkey subjects in three-dimensional structure from motion analysis. Further work is needed to determine the source of the difference between monkey and human subjects for the three-dimensional stimulus. Nonetheless the monkey and human subjects were similar among a wide range of other three-dimensional stimulus parameters suggesting that the ability to extract three-dimensional structure from motion is similar in monkey and man.

It is essential to point out that the monkey subjects may perform this task without actually perceiving the three-dimensionality of the stimuli. Two possible scenarios are that (1) the monkey is making comparisons between the edges of the display (where the velocity is

low) and the center of the display (where the velocity is high) and (2) the monkey is looking for uniformity in the velocity of the points in a local region of the display. The second possibility has been addressed above by the masking study, however, it is possible that a larger region than the 1.5 by 1.5° tested is needed to make such a measurement or that more points are needed to use this strategy. One result that argues against the monkeys using such cues is that they were all able to rapidly learn to generalize their behavior from the simpler rotation and expanding stimuli to the more difficult three-dimensional stimuli.

In the final analysis the above data indicate several similarities of monkey and human subjects on structure from motion tasks. However, it is very difficult to conclusively prove, with these data or any other data, that monkeys perceive three-dimensional structure from motion in the same way as human subjects. One could perhaps propose a task whereby the monkey performs a matching-to-sample task between objects described in motion and objects described by other visual submodalities (e.g., shape from shading, shape from disparity). However, even with such experiments, one still encounters the question of whether or not the animal is performing some high level matching task (e.g., small disparity differences against slow velocity) or actually *perceiving* three-dimensional structure. It may be that no matter how ingenuous a test is devised, the experimenter is ultimately limited by the use of behavioral data to demonstrate a cognitive function (Churchland, 1986; Edelman, 1990).

On the Differences in Neural Processing Required for the Rotation and Expansion versus the Three-Dimensional Tasks

The next question to be addressed is the different requirements to extract rotational or expansion motion versus three-dimensional structure from motion. For the subjects the major differences in their performance on these two sets of stimuli were

1. the number of points,
2. the point life,
3. spatial dependencies, and
4. reaction time in humans.

These differences probably reflect differences in the way in which the cortex may compute the two different percepts. The three-dimensional percept may require the prior construction of the two-dimensional flow fields. One can imagine that a cortical area first computes the underlying flow field for two-dimensional motion, which is sufficient to signal a two-dimensional field. Additional cortical areas are then needed to further process the signal to extract a solid three-dimensional structure. An anatomical and physiological homologue for this may be postulated to exist in the macaque monkey. Area MT is

thought to make local velocity field measurements (Allman et al., 1973; Zeki, 1974; Maunsell & Van Essen, 1987, 1983a). Area MT in the macaque projects to a secondary cortical structure known as MST (Maunsell & Van Essen, 1983b). Cells in area MST are sensitive to planar rotations and expansions and possibly three-dimensional rotation (Saito et al., 1986; Tanaka et al., 1986). As well neurons in area 7a appear to be involved in the analysis of structure from motion (Motter & Mountcastle, 1981; Sakata et al., 1984, 1985; Siegel, 1989). Other areas in the temporal visual path are also involved in extracting structure from motion. Cells in area TE (Perrett, 1985) located in the superior temporal sulcus respond to biological nonrigid motion (e.g., a walking human figure covered with lights; Johansson, 1982). It may be that in the three-dimensional structure from motion task, activity of cells in the dorsal motion path (MT/MST) needs to be integrated with the activity of cells in the ventral form pathways (e.g., TE) to form the complete percept of the three-dimensional form (in the current case—a cylinder). An anatomical basis for such integration between these areas exists. Area MST projects to area 7a and from there to the inferotemporal cortex (Andersen, 1986; Andersen et al., 1990). Thus it is possible that the inferotemporal cortex must be activated to perceive the three-dimensional shape of the stimulus. The longer point life dependence and the increased reaction time found with the cylinder stimulus may reflect the passage of action potentials through this multiplicity of visual areas.

A second possibility is that the greater “ease” of analysis of the rotation and expansion stimuli arises from the additional two-dimensional cues that are present in the expansion and rotation stimuli but not the three-dimensional stimulus. In the cylinder display only spatially distributed speed cues are available for detecting the difference between the unstructured and structured displays whereas in the expansion and rotation displays the direction as well as the speed of the dots are locally perturbed by decreasing the structure of the stimulus. For instance, an unstructured cylinder still has only the points moving left and right as is found in the structured case. However, the structured rotating stimulus will have motion trajectories locally moving in parallel whereas in the unstructured case they move in all directions. This “local parallel” motion cue is supported by the better performance in the masking experiments for the rotation and expansion stimuli as compared to the cylinder stimuli.

Implications for Computational Algorithms

As described in the Introduction, there are essentially two computational approaches to solving the three-dimensional structure from motion task: (1) flow and (2) spatial location. The result that a large number of points is required when the point life is short argues against a method based on the spatial location. If a spatial location

system was to be used, the visual system would have to perform the following process. It would need to compute which points in one presentation correspond to those in the next presentation. Then it would be necessary to directly determine if these frames corresponded to an underlying structured object. It is difficult to envision how this could be accomplished without implicitly computing motion velocity (cf. Grzywacz & Hildreth, 1987) in view of the large number of false matches and disappearing trajectories.

In a flow-based analysis, the visual system needs to first build up a map of velocity across the visual field. This velocity map is constrained by the temporospatial integrative properties of the motion-processing system (see below). This velocity map can then be processed by an additional visual region to determine if it corresponds to a three-dimensional structure.

This suggestion is also supported by physiological and lesion studies. The neurons that appear to extract structure from motion in monkey (Saito et al., 1986; Tanaka et al., 1986) are motion sensitive. Lesions of area MT, a known motion area, lead to deficits in tasks requiring motion analysis (Newsome, Wurtz, Dursteler, & Mikami, 1985; Wurtz, Dursteler, & Mikami, 1985; Siegel & Andersen, 1988; Newsome & Pare, 1988). As well preliminary evidence suggest that a lesion of area MST leads to short-term deficits in the ability to extract three-dimensional structure from motion with little effect on spatial acuity or simple motion perception (Siegel & Andersen, 1986). These results suggest therefore that it is flow and not position that is important to extract structure from motion. The question then arises as to the form of the algorithm that is implemented in cortex for the extraction of structure from motion. It is tempting to suggest that there is some form of simultaneous processing of many parts of the visual field as expressed in area MT (Siegel, 1988; Rodman & Albright, 1989).

The present data suggest that humans and monkeys have a similar ability to extract structure from motion under these controlled conditions, although there are still some open questions as to their differences. An essential underpinning of present-day neuroscience is that the neural mechanisms studied in these experimental situations would also be utilized in the natural environment. Thus it is the task of the visual neuroscientist to delineate the neural mechanisms underlying the structured motion analysis under highly controlled conditions and then to test the resultant model against natural visual input.

MATERIALS AND METHODS

We have defined the ability to extract structure from motion in terms of a reaction time task used previously in a study of shearing motion (Golomb et al., 1985). Both the human and monkey (*Macaca mulatta*) subjects performed the same task on the same equipment. The dis-

play screen was viewed from a distance of 57 cm so that 1 cm of movement on the screen corresponded to 1° of visual arc. The subjects controlled a small lever and were shaped³, or instructed, to the following task. The subject pulled back on the lever when the display came on. When the display changed in some way, the subject released the lever. The subjects had to pull the lever back initially within a time window of 150–800 msec. The change of the stimulus occurred after an interval of 1000–4000 msec of the onset of the display. Following the beginning of the change of the stimulus, the subject had to release the key within a time window. Different time windows for key release were used depending on the stimuli (see Results). If the subject completed the sequence of events correctly, then a reward was given (0.1 ml Tang orange juice for the monkeys; a click sound for the human subjects).

One can roughly compute the percent correct if the subject was releasing the key randomly (i.e., not detecting any change) as the ratio of the reaction time window to the total time needed for release (650 msec/3000 = 21%). Interestingly it was rare for subjects to have non-zero correct percentage hits for cases when there was no change from the control to the test stimulus. This can be attributed to the subject waiting for a change to occur. The trial would end before any change would occur; the subject would be left holding the key. Since subjects had no way of knowing in advance whether there was to be a change on that trial, they had no apparent motivation to release the key randomly. In some situations, however, subjects could tell that the task was more difficult (e.g., the presence of a cardboard mask over the display) and then they might change their strategy to one of random release and then end up with a number of trials correct when there was no change.

Description of the Motion Stimuli

Three different types of motion flow fields were used: (1) rotation in a plane, (2) expansion or compression in a plane, and (3) rotation in three dimensions (Fig. 1). Rotation in a plane is found naturally in situations when either the head of the observer is being turned about an axis parallel to the line of sight and/or an object is being turned about the same axis. Expansion in a plane is found when an observer is translated forward through space and/or an object is moved toward the observer. Compression is seen when the movements are in the opposite directions. (It should be noted that in these displays the radial velocity did not vary with the radial distance as is found naturally for constant velocity motion in depth.) An example of rotation in three dimensions would be the rotation about its vertical axis of a transparent beach ball with polka dots. For this last stimulus it is necessary to define a structured object to be rotated. In the present experiment a transparent hollow cylinder was used. The

velocity of a point at a position (X_i, Y_i) for each of these displays is given in Table 3.

The Controlled Structured and Unstructured Motion Stimuli

It was necessary to design a control unstructured stimulus so that the subject could detect a change in the reaction time task. To do this we shuffled the points in the display so as to keep the same motion components but remove their spatial relationships to each other. Thus the control unstructured display contained the same motion trajectories as the structured display except that the spatial structure of the velocity field was scrambled. The amount of structure in the display could be controlled by varying the degree of shuffling. We defined a display to have a fraction of structure of 1.0 when all the point trajectories were in their correct spatial locations. We defined a display to have a structure of 0.0 (i.e., to be unstructured) when each motion trajectory was randomly displaced by a random distance up to a limit of the width of the display. The trajectories of the points are

$$X_i(t+dt) = X_i(t) + dtV_x(X_i, Y_i, t)$$

$$X'_i(t+dt) = X_i(t+dt) + (1-F)WP \quad (1)$$

$$Y_i(t+dt) = Y_i(t) + dtV_y(X_i, Y_i, t)$$

$$Y'_i(t+dt) = Y_i(t+dt) + (1-F)WP$$

where $X_i(t)$, $Y_i(t)$ is the computed location of the i th point at t ; $X'_i(t)$, $Y'_i(t)$ are the displayed trajectories; dt is the time increment between displays; $V(X_i, Y_i, t)$ is the velocity of the point at a given position X_i, Y_i at time t ; F is the fraction structure of the display on the range [0,1]; W is the width of the display; and P is a number chosen from an even probability distribution over $[-0.5, 0.5]$. Note that $V(X_i, Y_i, t)$ defines the motion flow field. (Figure 2 demonstrates this shuffling for a planar rotating stimulus.)

Thus the task for the animal was to detect a change from an unstructured display (fraction of structure is 0) to a structured display (fraction structure is greater than 0 and less than or equal to 1). We went to great lengths to ensure that our reaction time test for structured motion could not be successfully completed by using other cues (e.g., point density, form). For example, in a stimulus of expansion, the point density becomes lower at the center of the display with time as all the points are moving away from the center. Furthermore, if one displays only the same set of points throughout a display, then it is possible for the subject to look for distortions in particular subsets of the display to perform the task.

To circumvent these two possibilities we drew on a method for displaying points with short lifetimes described by Zucker and Iverson (1986). The *point life* is defined as the amount of time a point is displayed before

Table 3. Equations for Velocity of Points for Three Different Flow Fields Used^a

Flow field used	Equation of motion
Rotation in a plane	$V_x = \omega\{X_i^2 + Y_i^2\}^{1/2} \sin(\omega t)$ $V_y = -\omega\{X_i^2 + Y_i^2\}^{1/2} \cos(\omega t)$
Expansion in a plane	$V_x = dr/dt X_i / \{X_i^2 + Y_i^2\}^{1/2}$ $V_y = dr/dt Y_i / \{X_i^2 + Y_i^2\}^{1/2}$
Rotation of hollow cylinder	$V_x = \omega\{X_i^2 + Z_i^2\}^{1/2} \sin(\omega t)$ $V_y = \dot{0}$

^a ω is the angular velocity of rotation of display; dr/dt is the rate of expansion of the display; Z_i is the location of the point in depth of a point on the surface of a cylinder; Z_i is computed as $\{R_c^2 - X_i^2\}^{1/2}$ where R_c is the radius of the cylinder.

disappearing. (Point life also has important computational significance, see Morgan & Ward, 1980; Zucker & Iverson, 1986.) The initial point life for each of the points, L_i is chosen on an even distribution on the range $[0, L]$, where L is the point life for the display. When the point's initial life time, L_i , has passed, the point vanishes from its present location and then reappears at a new location chosen at random from the surface of the screen. The point life, L_i , then becomes equal to the maximum life-time L . When the point has been displayed for its life time the process is repeated. Since each point is placed at a random position on the screen, points do not accumulate at locations such as the edges of the expansion display or edges of the cylinder.

If one were viewing a display with no motion, but with a point life of 200 msec, the screen would be full of points of light flickering on and off similar to the appearance of the sunlight on the surface of a lake in a light wind. Any possible static form cues in the image are constantly changing. Thus by using the point life, both of the problems of uneven point density and static form cues are avoided.

The use of short point life times removed an additional artifactual cue from the transition from unstructured to structured displays. This artifact would be due to the rapid change of all the points from their unstructured trajectories [$F = 0$ in Eq. (1)] to their structured trajectories ($F=1$). The velocity of each point at that time would have an added component unrelated to the structure of the display. This added velocity component at the time of change t_c can be computed from Eq. (1) as

$$V_x = [X_i(t_c + dt) - X_i(t_c)]/dt = V_x(X_i, Y_i, t_c) + (1-F)WP/dt$$

$$V_y = [Y_i(t_c + dt) - Y_i(t_c)]/dt = V_y(X_i, Y_i, t_c) + (1-F)WP/dt$$

The additional motion component $(1-F)WP/dt$ would be due to the rapid removal of the shuffling and would appear as a jump in the display.

To circumvent this problem, points were permitted to go from unstructured trajectories to structured trajectories only at the end of a point's path life. The only disadvantage to this strategy was that it took a time equal to the maximal point life L for all the points to go from their unstructured to their structured trajectories.

All of the human subjects had corrected 20/20 vision. The *M. mulatta* (3–6 years of age) were emmetropic as checked by streak retinoscopy. Monkey 60 had an opacity in the cornea of one eye that occurred at 2 years of age and thus viewed the display monocularly. The animal experiments were done in accordance with the statement guiding the care of use of animals approved by the Society for Neuroscience, April, 1984.

In summary, the displays consisted of flickering points that initially moved along trajectories selected from an underlying structure but displaced randomly on the screen. At some time, the point trajectories begin to appear at the correct spatial locations. Perceptually it appears that points are moving with no clear order and then changed to a clear moving structure.

Data Collection and Analysis

A Digital Equipment Company 11/73 computer controlled both the presentation of the visual stimuli as well as collection of the subject's performance. (The full details of stimulus generation and behavioral control are to be presented elsewhere, Siegel, in preparation.) A group of 10–120 different displays was run together. The displays consisted of movement of points whose initial positions were randomly selected on the screen. Each display had a different random number generator seed to ensure the subjects could not memorize a particular grouping of points. The displays were presented in fixed random block sequence. The subject viewed each display 30 times. The frame refresh rate of the displays was 35 Hz. In some experiments where the temporal properties of the motion system were examined, the displays were refreshed at 70 Hz as noted in the text. Intertrial intervals were 1–2.5 sec.

The subjects performance for each stimulus was tallied as (1) the percent of trials that the animal released the key within the reaction time window, (2) those trials where the animal released the key late, and (3) those trials when the key was released early. The mean and standard deviation of the reaction time of the correct trials for each stimulus were also computed. Individual trial results were not stored.

Plots were made of the percent correct trials as a function of the parameter varied in that run. In the cases where the data appeared to fit the error function, the logit function was used to compute the threshold with no correction made for the random percent hit rate (Ashton, 1972). Threshold was defined as the value of the dependent variable needed for the subject to get 50% of the trials correct. The asymptotic standard error, an estimate of the error of the threshold if an infinite number of measurements were made, was computed as in Ashton (1972). The goodness of the fit was evaluated by a χ^2 statistic. If the logit function was not a good fit (e.g., the dependence of the percent correct trials upon the varied parameter had a hump; $p > 0.1$), then linear interpolation was used to determine the 50% threshold over a range of values where the dependence of percentage correct was linear. An error estimate for the threshold was computed as the 95% confidence limits of the 50% threshold.

The stimuli in all cases consisted of moving points on a 25 cm diagonal oscilloscope screen (HP1311B, Phosphor P31). The points were drawn using a Data Translation DT2771 point plotting board so that each point was on for about 3 μ sec. Each stimulus presentation consisted of a fixed number of points moving with particular motion trajectories as described above.

Acknowledgements

This work was supported by grants from the NIH EY07492, the Sloan Foundation, the Whitaker Health Sciences Foundation, and a McKnight Foundation Scholar Award to R.A.A. Salary support for R.M.S. was provided by PHS NS07457 and the Charles H. Revson Foundation. Additional salary support for R.M.S. was provided from NIH EY05253 to Dr. Torsten N. Wiesel.

Reprint requests should be sent to Dr. Richard A. Andersen, Department of Brain and Cognitive Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

Notes

1. Current address: Ralph M. Siegel, Thomas J. Watson Research Laboratories, IBM, Route 134, Yorktown Heights, NY 10598.
2. Correspondence to Professor Richard A. Andersen, Department of Brain and Cognitive Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.
3. The monkeys were shaped by first training them on a dimming task and then allowing them to generalize to the more complex motion stimuli in a manner similar to that utilized by Golomb et al. (1985).

REFERENCES

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in the visual area MT of the macaque monkey. *Journal of Neurophysiology*, *51*, 16–31.

- Allman, J. M., Kaas, J. H., & Lane, R. H. (1973). The middle temporal visual area (MT) in the bushbaby, *Galago senegalensis*. *Brain Research*, *57*, 197–202.
- Andersen, R. A. (1986). The role of the inferior parietal lobule in spatial perception and visual-motor integration. In F. Plum, V. B. Mountcastle, & S. R. Geiger (Eds.), *Handbook of physiology. Section 1: The nervous system Volume V. Higher functions of the brain, Part 2*. Maryland: American Physiological Society. Pp. 483–518.
- Andersen, R. A., Asanuma, C., Essick, G. K. & Siegel, R. M. (1990). Cortico-cortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. *Journal of Comparative Neurology*, *296*, 65–113.
- Andersen, R. A., & Siegel, R. M. (1986). Two and three dimensional structure from motion sensitivity in monkey and humans. *Society for Neuroscience Abstracts*, *10*, 476.
- Andersen, R. A., & Siegel, R. M. (1989). Motion processing in primate cortex. In G. M. Edelman, W. E. Gall and W. M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps*. New York: Wiley.
- Ashton, W. D. (1972). In A. Stuart (Ed.), *The logit transformation with special reference to its uses in bioassay*. London: Charles Griffin and Co. Ltd.
- Baker, J. F., Petersen, S. E., Newsome, W. T., & Allman, J. M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*), a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. *Journal of Neurophysiology*, *45*, 397–416.
- Bennett, B. M., & Hoffman, D. D. (1985). The computation of structure from fixed-axis, motion, nonrigid structures. *Biological Cybernetics*, *51*, 293–300.
- Bertrand, M. (1969). *The behavioral repertoire of the stump-tail macaque: A descriptive and comparative study*. Basel: S. Karger.
- Beverly, K. I., & Regan, D. (1985). The relation between discrimination and sensitivity in the perception of motion in depth. *Journal of Physiology (London)*, *249*, 387–398.
- Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B*, *290*, 137–151.
- Cornilleau-Peres, V., & Droulez, J. (1987). Visual perception of surface curvature: from computational analysis to psychophysical experiments. *Neuroscience*, *22*, S843.
- Churchland, P. S. (1986). *Neurophilosophy*. Cambridge: MIT Press.
- Desimone, R., & Ungerleider, L.G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the Macaque. *Journal of Comparative Neurology*, *248*, 164–189.
- Edelman, G. M. (1990). *The remembered present*. New York: Basic Books.
- Golomb, B., Andersen, R. A., Nakayama, K., MacLeod, D. I. A., & Wong, A. (1985). Visual thresholds for shearing motion in monkey and man. *Vision Research*, *25*, 813–820.
- Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity based formulations. *Journal of the Optical Society of America*, *4*, 503–518.
- 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 W. von Seelen, G. Shaw & B. Leinos (Eds.), *Organization of neural networks*. U. M. Wienheim: VCH Press. Pp. 239–259.
- Hildreth, E. C. (1983). *The measurement of visual motion*. Cambridge: MIT Press.
- Hildreth, E. C. (1984). The computation of the velocity field.

- Proceeding of the Royal Society of London B*, 221, 189–220.
- Hubel, D., & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proceeding of the Royal Society of London B*, 198, 1–59.
- Johansson, G. (1982). Visual space perception through motion. In A. H. Wertheim, W. A. Wagenaar, & H. W. Leibowitz (Eds.), *Tutorial on motion perception*. New York: Plenum Press.
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceeding of the Royal Society of London B*, 208, 385–397.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- Maunsell, J. H. R., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49, 1127–1147.
- Maunsell, J. H. R., & Van Essen, D. C. (1983b). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3, 2563–2586.
- Maunsell, J. H. R., & Van Essen, D. C. (1987). The topographic organization of the middle temporal visual area in the macaque monkey and its relationship to figure-ground perception. *Journal of Neuroscience*, 7, 943–958.
- Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *Journal of Neurophysiology*, 55, 1308–1327.
- Morgan, M. J., & Ward, R. (1980). Interocular delay produces depth in subjectively moving noise patterns. *Quarterly Journal Experimental Psychology*, 32, 387–95.
- Motter, B. C., & Mountcastle, V. B. (1981). The functional properties of the light-sensitive neurons of the posterior parietal cortex studied in waking monkeys: Foveal sparing and opponent vector organization. *Journal of Neuroscience*, 1, 3–26.
- Nakayama, K., & Tyler, C. W. (1981). Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research*, 21, 427–433.
- Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8, 2201–2211.
- Newsome, W. T., Wurtz, R. H., Dursteler, M. R., & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of extrastriate area MT of the macaque monkey. *Journal of Neuroscience*, 5, 824–840.
- Perrett, D. I., Smith, P. A., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D., & Jeeves, M. A. (1985). Visual analysis of body movements by neurones in the temporal cortex of the macaque monkey: A preliminary report. *Behavioral and Brain Research*, 16, 153–170.
- Petersen, S. E., Baker, J. F., & Allman, J. M. (1980). Dimensional selectivity of neurons in the dorsolateral visual area of the owl monkey. *Brain Research*, 197, 507–511.
- Rodman, H. R., & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, 75, 53–64.
- Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6, 145–157.
- Sakata, H., Shubutani, H., Ito, Y., & Tsurugai, K. (1984). Parietal visual neurones respond to rotary movements of the visual stimulus. *Journal of Physiological Society of Japan*, 46, 45.
- Sakata, H., Shibutani, H., Kawano, K., & Harrington, T. L. (1985). Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Research*, 25, 453–463.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Siegel, R. M. (1988). Discovery of structure from motion in monkey, man and machine. In D. Z. Anderson (Ed.), *Neural information processing systems*. New York: American Institute of Physics.
- Siegel, R. M. (1989). Motion analysis in the inferior parietal area 7a of the rhesus monkey. *Society of Neuroscience Abstracts*, 15, 162.
- Siegel, R. M., & Andersen, R. A. (1988). Perception of three-dimensional structure from visual motion in monkey and man. *Nature (London)*, 331, 259–261.
- Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, 6, 134–144.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.
- Ungerleider, L. G., & Desimone, R. (1986a). Cortical connections of visual area MT in the macaque. *Journal of Comparative Neurology*, 248, 190–222.
- Ungerleider, L. G., & Desimone, R. (1986b). Cortical connections of visual area MT in the Macaque. *Neuroscience Letters*, 67, 153–158.
- Van Essen, D. C., Maunsell, J. H. R., & Bixby, J. L. (1981). The middle temporal visual area in the macaque: Myeloarchitecture, connections, functional properties and topographic organization. *Journal of Comparative Neurology*, 199, 293–326.
- von Helmholtz, H. (1910). In J. P. C. Southall (Ed.), *Treatise on physiological optics*. New York: Dover Publications.
- Wallach H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45, 205–217.
- Wertheim, A. H., Wagenaar, W. A., & Leibowitz, H. W. (1982). *Tutorial on motion perception*. New York: Plenum Press.
- Wurtz, R. H., Dursteler, M. R., & Mikami, A. (1985). Deficits in visual motion processing following ibotenic acid lesions of the middle temporal visual area of the macaque monkey. *Journal of Neuroscience*, 5, 825–840.
- Zeki, S. (1980). The response properties of cells in the middle temporal area (area MT) of owl monkey visual cortex. *Proceeding of the Royal Society of London B*, 207, 239–248.
- Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology (London)*, 236, 549–573.
- Zucker, S. W. (1983). The fox and the forest: Towards a type I/type II constraint for early optical flow. Computer Vision and Robotics Laboratory, TR-83-11, McGill University.
- Zucker, S. W. (1985). Early orientation selection: Tangent fields and the dimensionality of their support. Computer Vision and Robotics Laboratory, TR-85-13-R, McGill University.
- Zucker, S. W., & Iverson, L. (1986). From orientation to virtual lines. *Biological Cybernetics*, 54, 407–416.