

# Engagement of Gaze in Capturing Targets for Future Sequential Manual Actions

YASUO TERAO,<sup>1</sup> N. E. MICAEL ANDERSSON,<sup>1</sup> J. RANDALL FLANAGAN,<sup>2</sup> AND ROLAND S. JOHANSSON<sup>1</sup>

<sup>1</sup>*Section for Physiology, Department of Integrative Medical Biology, Umeå University, SE-901 87 Umeå, Sweden; and*

<sup>2</sup>*Department of Psychology, Canadian Institutes of Health Research Group in Sensory-Motor Systems and Centre for Neuroscience Studies, Queen's University, Kingston, Ontario K7L 3N6, Canada*

Received 11 February 2002; accepted in final form 13 June 2002.

**Terao, Yasuo, N. E. Micael Andersson, J. Randall Flanagan, and Roland S. Johansson.** Engagement of gaze in capturing targets for future sequential manual actions. *J Neurophysiol* 88: 1716–1725, 2002; 10.1152/jn.00101.2002. We investigated the role of saccadic gaze fixations in encoding target locations for planning a future manual task consisting of a sequence of discrete target-oriented actions. We hypothesized that fixations of the individual targets are necessary for accurate encoding of target locations and that there is a transfer of sequence information from visual encoding to manual recall. Subjects viewed four targets presented at random positions on a screen. After various delays following target extinction, the subjects marked the remembered target locations on the screen with the tip of a hand-held stick. When the targets were presented simultaneously among distracting elements, the overall accuracy of marking increased with presentation time and total number of targets fixated because the subjects had to serially fixate the individual targets to locate them. Without distractors, the marking accuracy was similarly high regardless of duration of target presentation (0.25–8 s) and number of targets fixated; it was comparable to that with distractors when all four targets had been fixated. This indicates parallel encoding of target locations largely based on peripheral vision. Location memory was stable in these tasks over the delay periods investigated (0.5–8 s). With parallel encoding there was a “shrinkage” in the visuomotor transformation, i.e., the distances between the markings were systematically smaller than the corresponding inter-target distances. When the targets were presented sequentially without distractors, marking accuracy improved with the total number of targets fixated and shrinkage in the visuomotor transformation occurred only with parallel encoding, i.e., when subjects did not fixate the targets. In all experimental conditions for trials in which targets were fixated during encoding, there was little correspondence between the marking sequence and the sequence in which the targets were fixated. We conclude that subjects benefit from fixating targets for subsequent target-oriented manual actions when the targets are presented among distractors and when presented sequentially; when distinct targets are presented simultaneously against a blank background, they are efficiently encoded in parallel largely by peripheral vision.

## INTRODUCTION

In natural activities, subjects control gaze shifts and fixations proactively to gather spatial information for planning and control of subsequent actions. Proactive control of gaze has been

shown in driving (Land 1992; Land and Horwood 1995; Land and Lee 1994), music reading (Goolsby 1994; Kinsler and Carpenter 1995; Land and Furneaux 1997), typing (Inhoff and Wang 1992), walking (Patla and Vickers 1997), throwing in basketball (Vickers 1996), putting in golf (Vickers 1992), and batting in cricket (Land and McLeod 2000). In manipulatory tasks, gaze fixations support the planning of hand movements by marking key positions to which the fingertips or grasped object are directed (Ballard et al. 1992; Johansson et al. 2001; Land et al. 1999; Smeets et al. 1996; see also Abrams et al. 1990; Binsted and Elliott 1999; Bock 1986; Neggers and Bekkering 2000; Pélissier et al. 1986; Prablanc et al. 1986). Natural manipulatory tasks usually consist of a series of phases for which gaze fixations of critical landmarks—through retinal and extraretinal information—provide spatial reference points (Johansson et al. 2001; Land et al. 1999). Therefore memory could be useful to buffer visually acquired spatial information across successive action phases. Ballard et al. (1992) examined eye-hand coordination when subjects arranged a series of colored blocks to match a visible model. They concluded that subjects used memory of the model to guide the hand actions but for no more than one or two subsequent action phases. Ballard suggested that subjects prefer to repeatedly view the model to avoid overloading working memory. However, memory may play a greater role in many natural situations in which we are engaged in multiple tasks in visually complex environments (Kowler 1995). For instance, when we reach for a spoon, take sugar from a basin, and put it in our morning coffee while reading the newspaper, we largely rely on memory and/or peripheral vision to guide our manipulatory actions.

In the present study, we investigated the role of saccadic gaze fixations in extracting information from a scene for planning a future manual task consisting of a sequence of discrete manipulatory actions. Subjects viewed a display including four targets and then, after the targets were extinguished, used a hand-held stick to mark the remembered locations of the targets on the same display. Thus the subjects had to encode the locations of the targets in memory while exploring the scene and then had to recall these locations to perform the task. In natural environments, objects toward which manipulatory actions are directed may be visually salient and located primarily by peripheral vision whereas in other instances, visual search may be required for target detection based on central vision. To experimentally address these two conditions, we examined two

Address for reprint requests: Y. Terao, Dept. of Neurology, Division of Neuroscience, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113-8655 (E-mail: yasuo.terao@physiol.umu.se or yterao-tyk@umin.ac.jp).

types of scenes: one in which the four targets were presented against a blank background and one in which they were presented among distractors. With distractors, subjects had to search for and fixate the targets, whereas, without distractors, the targets could be detected by peripheral vision. In addition, we also manipulated the time of target presentation. With the shortest presentation, hardly any eye movements could be made prior to target extinction and encoding of target localization relied largely on peripheral vision. With longer presentations, subjects had enough time to fixate the targets. We tested the hypothesis that target fixation is necessary for optimum encoding of target location. That is, we predicted that the accuracy of manual performance would increase with the number of targets fixated. We also predicted that subjects would strive to fixate all targets. Second, given that memory for a scene may not only contain spatial features but also information about the gaze sequence used while capturing the scene (Noton and Stark 1971a,b), we also investigated whether there is a transfer of sequence information from visual encoding to manual recall. That is, we tested the hypothesis that the sequence by which the remembered locations of the targets are marked reflects the order of fixations used while visually capturing the targets.

## METHODS

### Subjects and general procedure

Sixteen subjects participated in the present experiments after providing informed consent. The experimental protocol was conducted according to the declaration of Helsinki. None of the subjects required corrective lenses or had a history of ophthalmological or neurological disease. While seated behind a table, the subject could see a computer screen (28 × 21 cm, Sharp Color TFT-LCD Module, Tokyo, Japan, with SVGA interface) located on a horizontal support surface formed by the top of a wooden stand placed on the table. The screen was aligned in a frontal plane, termed the work-plane, and the center of the screen was located 45 cm straight in front of the subject's eyes.

In each trial, four white targets were presented for a specified duration against a black background (Fig. 1A). During this encoding period, the targets were presented either simultaneously or sequentially at randomly selected locations on the screen. The choice of four targets was based on Treisman's suggestion (1999) that in studies of working memory, the attentional limit is around four elements during perception of brief displays. Furthermore, observers can enumerate up to four objects rapidly and accurately, whereas greater numbers take far longer and are enumerated less accurately (Pylyshyn 2000). Following a delay period after the targets were extinguished, the subjects used a stick, held by the preferred hand, to mark the remembered target locations on the blank screen (recall period).

### Apparatus

The stick used to mark the targets was 15 cm long and 1.3 cm in diameter. The distal 6.0 cm of the stick was conical, making it pointed. The three-dimensional position of the tip of the stick was recorded at 60 samples/s using a miniature electromagnetic position-angle sensor (FASTRAK, Polhemus, Colchester, VT) attached at the proximal end of the stick. In the experimental environment, the accuracy of the position measurements was  $\pm 0.2$  cm in the plane of the screen. To reduce electromagnetic interference, the metal frame of the TFT-LCD screen had been removed.

The apparatus for gaze recording has been described previously (Johansson et al. 2001). Briefly, we used an infrared video-based eye-tracking system (RK-726PCI pupil/corneal tracking system,

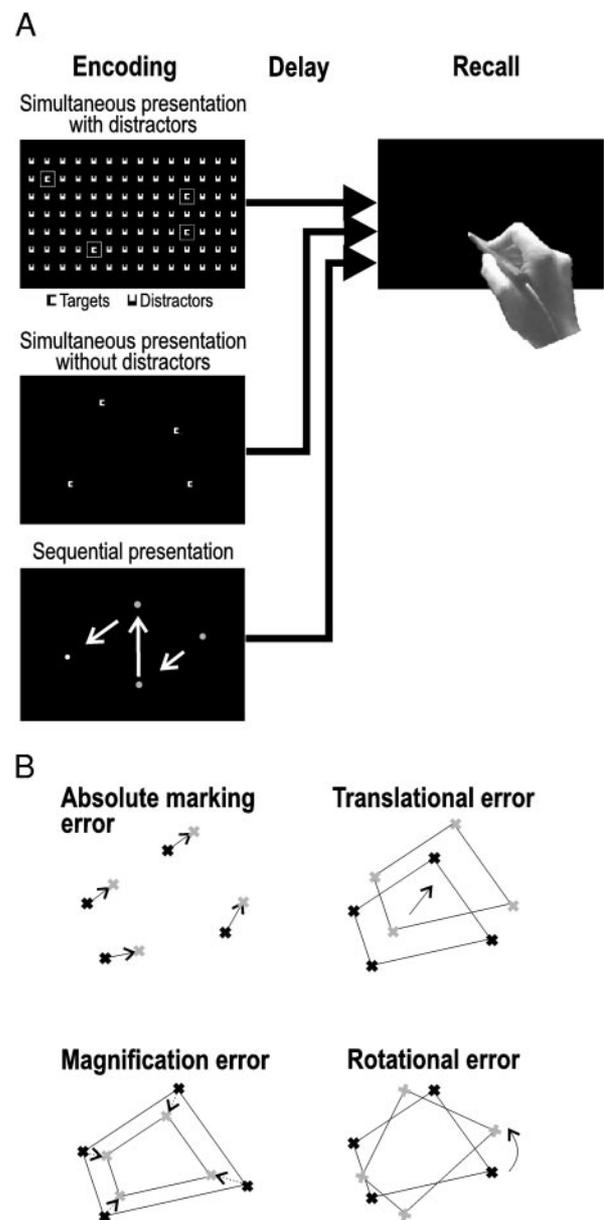


FIG. 1. Tasks and assessment of performance. *A*: each of the 3 different types of target presentations (simultaneous presentation of targets with and without distractors and sequential target presentation) were followed by a delay after which the subjects marked the remembered locations of the targets on the display using the tip of a hand-held stylus. In the sequential target presentation the 4 targets (○) appeared sequentially on the blank screen. *B*: schematic illustrations of the error measurements used to assess manual performance.

Iscan, Burlington, MA) to record the position of gaze of the right eye in the plane of the screen at 120 samples/s; the view of the left eye was always blocked. The eye-imaging camera together with the infrared light source and the dichroic mirror were mounted on a wooden frame that was fixed to the table. To stabilize the head, subjects bit on a U-shaped stainless steel plate anchored to the support frame of the apparatus. Both sides of the plate were coated with dental wax, and the head was effectively stabilized by impressions of the dentition made in the wax prior to gaze recording. We used a two-step calibration procedure to obtain gaze data with satisfactory spatial accuracy (Johansson et al. 2001). An initial calibration using Iscan's "Line-of-Site Plane Intersection Software" was followed by calibration measurements repeatedly taken during the experiments using a nine-point

calibration array. The nine calibration points (3 rows and 3 columns) covered an  $18 \times 12$ -cm (width  $\times$  height) area centered on the screen; the targets were always located within this area. In each calibration, the nine points were measured twice, and calibration measurements were taken every 10–30 trials. Each sampled data point during the experiment was calibrated off-line using data obtained from the nearest calibration measurements before and after the point. A satisfactory gaze recording required fixing the subject's eyebrow in an uplifted position by attaching tape between the eyebrow and the forehead in a manner that did not prevent the subject from blinking. The standard deviations of the error distributions of gaze position measurements in the horizontal and vertical dimensions were  $0.50$  and  $0.52^\circ$  angle of gaze (see Johansson et al. 2001 for further details). This corresponds to  $0.39$  and  $0.41$  cm distance in the work plane, respectively.

A PC (microcomputer) was used to control the experiments (target presentations, delay periods, etc.). Data were sampled and analyzed using the SC/ZOOM system (Physiology Section, IMB, Umeå University). Gaze and kinematic data were time-synchronized and stored at 200 Hz using linear interpolation between consecutive measurements.

### Tasks

**ENCODING AND RECALL.** *Target presentation—encoding.* Three types of displays were employed for target presentation (Fig. 1A): simultaneous presentation with targets embedded among distractors, simultaneous presentation without distractors, and sequential presentation without distractors. For each display, center positions of the four targets were randomly selected from an imaginary orthogonal grid array (14 vertical columns  $\times$  8 horizontal rows) where adjacent grid points were separated by 13.6 mm in the horizontal direction and 16.4 mm in the vertical direction. Before and after the target presentation, the monitor screen was blank.

For *simultaneous presentation with distractors*, all four targets were white with a C-shaped form, and the distractors, located at the remaining positions of the imaginary grid, were U-shaped (Fig. 1A). Both the targets and the distractors were 3.4 mm ( $0.43^\circ$ ) in width and 4.1 mm ( $0.52^\circ$ ) in height. Central vision was required to discriminate the targets from the distractors (see RESULTS). During *simultaneous presentation without distractors*, the target positions were selected randomly and the remaining grid positions were blank. During *sequential target presentation*, the four targets appeared sequentially at random locations on the blank screen; each target disappeared when the next appeared. The targets were white circles 3.4 mm in diameter ( $0.43^\circ$ ), and there were no distractors.

*Hand actions—recall.* After a delay period following the end of the target presentation, an auditory cue (1,000-Hz tone for 150 ms) instructed the subjects to start marking each of the remembered target locations on the blank screen. The subjects were instructed to lift the stick from the screen between consecutive markings rather than to slide it over the screen. The coordinates of the tip of the stick when it approached the screen within a distance of 1.5 mm were taken as the marked position of a target. The subjects were instructed to always make four markings even if they were uncertain about the target locations and were encouraged to be as precise as possible. No instruction was given as to the sequence in which the four locations were to be marked. The trial ended after four positions had been marked. At that point, the subject received feedback about their performance; the mean absolute distance between the actual and marked target locations was numerically displayed on the screen for 0.5 s (see *Data analysis*). The subsequent trial commenced 1–2 s after the feedback was extinguished and the screen went blank. In all tasks, subjects were free to move their eyes as they wanted. In control experiments, we assessed the intrinsic inaccuracy of the marking process by having the subjects perform the marking task with visible targets. The marking error measured as the straight distance between the recorded location of the tip of the stick and the center of the target

was  $0.34 \pm 0.03$  (SD) cm for the display without distractors and  $0.34 \pm 0.02$  cm for the display with distractors.

**TEST SERIES.** Ten subjects (6 males, 4 females, age 20–31 years; 8 right-handed, 2 left-handed) performed one test series with simultaneous presentation of targets with distractors and one series without distractors. In each series, the targets were presented for 10 different durations (0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0 s). There were 10 trials for each of these durations for a total of 100 trials, and the various durations occurred in an unpredictable sequence. The delay period between target presentation and the auditory cue that instructed the subjects to start marking the target locations was fixed at 0.5 s. To assess the possible influence of the delay period on the marking performance, six different subjects (3 males, 3 females, age 19–33 years; 5 right-handed and 1 left-handed) participated in separate experiments with simultaneously presented targets (with and without distractors) where nine different delay periods were used (0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0 and 8.0 s). There were 10 trials for each of these durations for a total of 90 trials, and the various delays occurred in an unpredictable sequence. The duration of target presentation was fixed at 8 s.

The same six subjects participated in the test series with sequential target presentation. Each of the four sequentially appearing targets stayed on the screen for 0.2, 0.5, 1.0, or 2.0 s in different blocks of trials with the entire encoding period lasting for 0.8, 2, 4, or 8 s, respectively. Each block consisted of 40 trials.

The average center-to-center distances between all pairs of targets during simultaneous presentation with and without distractors and sequential presentation were  $8.1 \pm 4.0$ ,  $8.3 \pm 4.0$ , and  $8.1 \pm 4.0$  cm, respectively. The corresponding median values were 8.0, 8.1, and 7.8 cm. For all test series, the distances ranged from 1.4 to 21.0 cm. Thus the separation of the targets was on average about  $10^\circ$  visual angle and the most eccentric targets were located  $13.4^\circ$  from the center of the screen. With this rather limited target separation, the head-fixed condition in the present experiments should not have compromised appreciably the subjects marking performance (e.g., Biguer et al. 1984). The sequence of the test series performed by each group of subjects was counterbalanced across the subjects involved.

### Data analysis

**GAZE MEASUREMENTS.** We measured the positions of gaze fixations during the encoding period. The onset and end of each fixation were defined as the times when gaze velocity (low-passed filtered at 30 Hz with a second-order Butterworth filter) decreased below and exceeded 15 cm/s ( $19.1^\circ$ /s), respectively. Fixation of a target was deemed to have occurred when gaze stayed within a radius of 2 cm ( $2.6^\circ$ ) from the center of a target for at least 50 ms (see RESULTS).

**MEASURES OF MANUAL PERFORMANCE.** We first determined which of the four marked locations most likely corresponded to the four targets by finding the combination that was associated with the minimum sum of squared distances between the targets and the marked locations. As a measure of manual performance, we computed the *absolute marking error* by taking the mean value across the four targets of the straight distance between the location of the center of the target and the corresponding marked location. The error was further decomposed into three components: *translational error*, *magnification error*, and *rotational error* (Fig. 1B). We defined the translational error as the vectorial distance between the “center of gravity” for the target locations and that of the corresponding marked locations. The location of the center of gravity was defined by the mean  $x$  and  $y$  coordinates of the relevant locations. The magnification factor is an error measure that refers to the “expansion” or “shrinkage” of the shape defined by the four marked locations with reference to the corresponding shape defined by the target locations. To estimate this factor, we first computed, for each of the six inter-target distances

given by the four targets, the ratio of the distance between the marked locations and the corresponding target locations. The magnification factor was then defined as the mean ratio for the six inter-target distances. A factor greater than unity indicates an overall expansion of the marked image, whereas a factor less than unity indicates shrinkage. The rotational error was defined as the average angular error between the marked locations and the corresponding target locations where the center of gravity of the target locations defined the origin of rotation. The rotational error was computed after the center of gravity of the marked image had been translated to that of the target image and then scaled by the inverse of the magnification factor. The percentages of absolute marking error attributed to the translation, magnification, and rotational errors corresponded to their contribution to the absolute marking error, respectively, as scored during this stepwise procedure. That is, the order by which these error measures were computed would influence the estimate of their contribution, with the exception of the magnification factor. The error that remained after translating, magnifying, and rotating the marked locations is referred to as the *residual error*.

**TRANSFER OF SEQUENTIAL INFORMATION BETWEEN ENCODING AND RECALL.** For simultaneous presentation of targets among distractors, we analyzed whether the sequence in which the gaze located the targets was related to the sequence in which the targets were marked. We restricted the analysis to trials in which the subjects had foveated all targets [fixations within a distance of 2 cm ( $2.6^\circ$ ) from the center of each target]. We likewise excluded trials in which a fixation simultaneously captured two targets, i.e., the two targets were within 2 cm of a fixation point. For sequential presentation, we analyzed whether the sequence in which the targets were presented was related to the sequence in which they were marked. Again, we restricted the analysis to trials in which the subjects had foveated all targets.

For each trial, the targets were labeled 1–4 according to the temporal sequence in which they were first fixated during simultaneous presentation and presented (and fixated) during the sequential presentation. For the simultaneous presentation, we also scored the sequence with which the four targets were last fixated, i.e., it was noted that targets were sometimes fixated more than once. The order in which the four targets were marked during recall was then represented by a sequence of four digits referring to the target labels. For instance, if the order in which a subject marked the targets was the same as that with which the targets were fixated, the resulting sequence was 1 2 3 4. We separately analyzed the sequence of the first four fixations and the sequence of last four fixations. To test whether there was a transfer of sequence information from encoding to recall, using standard statistical procedures, we calculated the upper 95% confidence limit for the probability of occurrence of this sequence postulating that order of marking would have been random; an observed frequency above this limit was taken as an evidence for the transfer of sequence information.

**STATISTICAL ASSESSMENT.** We used repeated-measures ANOVAs to assess the effects of the mode of target presentation, duration of presentation, and the delay between presentation and recall on error measures of manual performance as well as measures of gaze behavior. The effect of target fixation on the error measures of manual performance was also assessed by repeated measures ANOVA.

## RESULTS

### Manual performance after simultaneous presentation of targets with and without distractors

Figure 2A shows the absolute marking error as a function of the presentation time for simultaneous target presentation with and without distractors. This error represented the average distance between the target locations and the corresponding remembered locations marked on the screen. Without distrac-

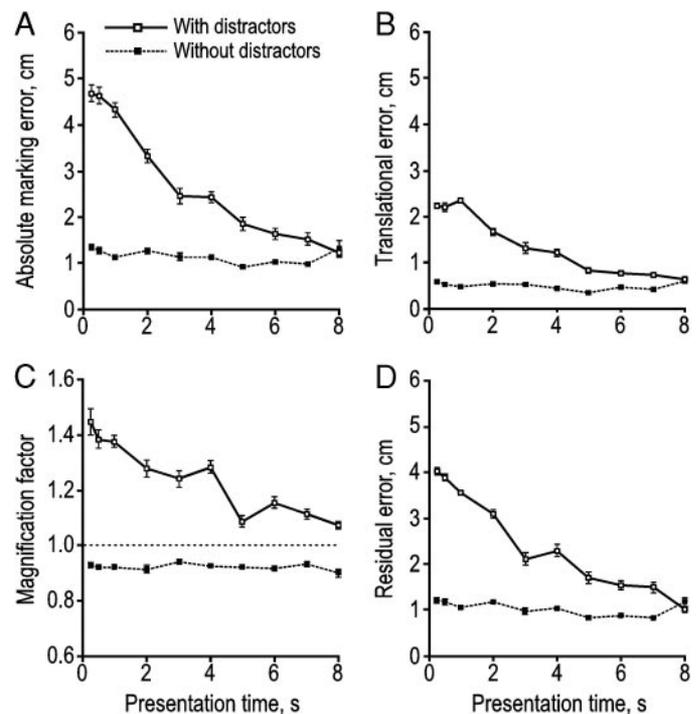


FIG. 2. Manual performance after simultaneous target presentation with and without distractors. The absolute making error (A), translational error (B), magnification factor (C), and residual error (D) were plotted as a function of the target presentation time. Symbols indicate mean values and error bars indicate SE for individual subjects.

tors, the absolute error did not vary with target presentation time [ $F(9,81) = 1.39, P = 0.19$ ]. The mean error was 1.16 cm. In contrast, when the targets were presented among distractors, the error was influenced by the presentation time [ $F(9,81) = 49.76, P < 0.001$ ]. The error was greatest (mean: 4.68 cm) at the shortest presentation time (0.25 s; chance performance was estimated at  $4.66 \pm 0.13$  cm) and decreased with increasing presentation time. At the 8-s target presentation time, the error (1.23 cm) was comparable to that observed without distractors (Student's *t*-test:  $P = 0.69$ ). These results indicate that subjects needed to fixate the targets to detect and encode their locations when the targets were presented among distractors, but not when they were presented without distractors.

**TYPES OF ERROR IN THE VISUOMOTOR TRANSFORMATION.** For simultaneous presentation without distractors, the translational, magnification and rotational errors explained 17.0, 18.6, and 1.7% of the absolute marking error, respectively (mean across all the presentation times). With distractors, the translational, magnification, and rotational errors accounted for 16.4, 7.9, and 3.9 percent of the error.

Figure 2, B–D, plots the translational error, the magnification factor, and the residual error for the same data shown in Fig. 2A. With distractors, all these error measures decreased reliably with presentation time [translational error,  $F(9,81) = 27.90, P < 0.001$ ; magnification factor,  $F(9,81) = 5.12, P < 0.001$ ; residual error,  $F(9,81) = 28.19, P < 0.001$ ]. In contrast, without distractors (---), the translational and magnification errors did not change with the presentation time [translational error,  $F(9,81) = 1.00, P = 0.44$ ; magnification factor,  $F(9,81) = 0.37, P = 0.95$ ]. The presentation time influenced

the residual error [ $F(9,81) = 2.22, P = 0.03$ ], but the effect was small compared to the presentation with distractors.

On average, the magnification factor was greater than unity for simultaneous presentation with distractors and below unity for simultaneous presentation without distractors irrespective of the exposure time. Without distractors, the magnification factor was 0.92 on average and reliably less than unity (Student's  $t$ -test:  $P < 0.001$ ). Even at the longest presentation time (8 s), the magnification factor for simultaneous presentation with distractors (mean: 1.07) was significantly greater than the magnification factor observed without distractors (Student's  $t$ -test:  $P < 0.001$ ).

The rotational error did not vary with the mode of presentation [ $F(1,9) = 0.45, P = 0.50$ ] or with the target presentation time [ $F(9,81) = 0.82, P = 0.60$ ] nor was there any interaction between these factors [ $F(9,81) = 0.59, P = 0.81$ ]. Overall, the rotational error did not differ from 0 degrees (Student's  $t$ -test:  $P = 0.06$ ; range:  $-5.34$ – $3.27^\circ$ ).

In summary, for the longest target presentation (8 s) when the subjects should have been able to fixate all the targets during the encoding period, all the error measures except for the magnification factor were comparable for both types of presentation. An expansion of the shape defined by the four targets was observed when marking targets presented simultaneously with distractors, whereas shrinkage was observed when marking targets presented without distractors. Because the targets were relatively accurately located, the marking error was not due to the target not being encoded.

**TARGET FIXATIONS DURING THE ENCODING PERIOD.** To assess whether a target was fixated during the encoding period, we first needed to estimate the size of the "functional fovea"—the variation in gaze positions observed when subjects fixate a given target. For each target, we analyzed the absolute marking error as a function of the distance between the target and its closest fixation point during target presentation. These distances were sorted in 1-cm bins for this purpose and data obtained for all presentation times (0.25–8 s) were pooled. We focused on simultaneously presented targets and separately analyzed targets presented with and without distractors (Fig. 3). For simultaneous presentation with distractors, the absolute marking error was small for distances less than 2 cm, averaging 1.61 cm. The error depended on the distance [ $F(9,81) = 6.09, P < 0.001$ ] and increased markedly as the distance exceeded 2 cm (Fig. 3, —; only distance less than 10 cm were considered for statistics because there were few trials with longer distances). Thus the effective region for target detection was within a radius of 2 cm ( $2.6^\circ$ ) from the fixation point, and we considered a target to be fixated if it was located within this radius from the measured point of gaze. The gradual increase in marking error for distances greater than 2 cm was most probably due to a gradual decrease in the likelihood of target detection. For targets presented simultaneously without distractors (Fig. 3, - - -), the absolute marking error did not depend on the gaze-to-target distance [ $F(9,81) = 0.58, P = 0.81$ ].

For each trial, we counted the number of targets fixated at least once and computed the fractions of trials in which 0, 1, 2, 3, or 4 targets were fixated. For simultaneous presentation with distractors, the subjects could not fixate any of the targets at the shortest presentation time (0.25 s). The number of targets

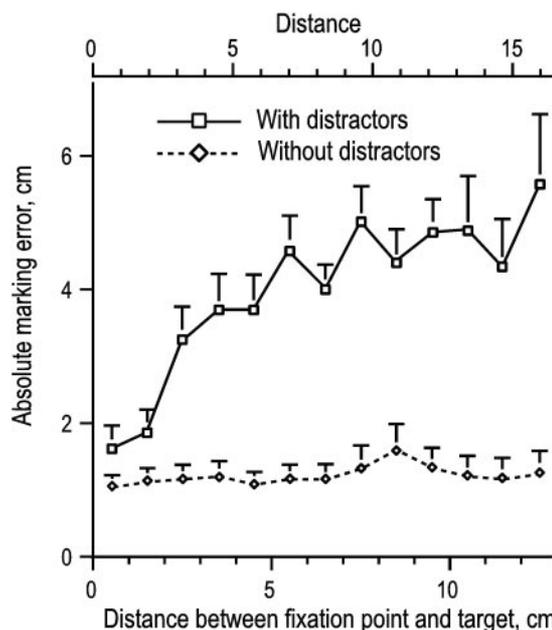


FIG. 3. The absolute marking error is shown as a function of the distance between a target and its nearest fixation point during the encoding period (data grouped in 1-cm distance bins). For each trial, the marking errors for the 4 targets were analyzed separately and pooled (the distance between a target and its nearest fixation point and the marking error was recorded separately for the 4 targets in each trial). The symbols indicate mean values and the error bars indicate SE. — and - - -, data obtained during simultaneous target presentation with and without distractors, respectively. *Bottom* and *top abscissas* represent measurements scaled in distance on the work plane and in degrees of visual angle. With distractors, note the steep increase in marking error between 2 and  $3^\circ$  visual angle.

fixated increased with the time of presentation (Fig. 4A, *top*). At the longest target presentation time (8 s), the subjects fixated all four targets in 77.5% of the trials and three targets in 12.5% of the trials. For simultaneous target presentation without distractors, the number of targets fixated also increased with the time of presentation (Fig. 4B). However, even at the longest target presentation (8 s), the subjects fixated all targets in only 40% of the trials, whereas they fixated two and three targets in 25 and 22.5% of the trials, respectively. Thus the subjects were less prone to fixate all targets in the absence of distractors.

**RELATIONSHIP BETWEEN MARKING ACCURACY AND TARGET FIXATIONS.** The *lower three panels* in Fig. 4A and B relate the manual performance for simultaneous target presentation with and without distractors to the total number of targets fixated during the presentation period. Without distractors (Fig. 4B), the number of targets fixated influenced neither the absolute marking error averaged across all four targets [ $F(4,36) = 1.06, P = 0.39$ ] nor the magnification factor [ $F(4,36) = 1.18, P = 0.34$ ]. The residual error decreased slightly with the number of fixated targets [ $F(4,36) = 3.39, P < 0.02$ ]. In contrast, when the targets were presented among distractors (Fig. 4A), the absolute marking error was markedly influenced by the number of targets fixated [ $F(4,36) = 28.92, P < 0.0001$ ]. The error decreased with the number of targets fixated. When none of the targets was fixated, the mean error ( $5.04 \pm 0.12$  cm) was approximately at chance level, whereas for trials in which all four targets were fixated, the error was comparable to that during simultaneous presentation without distractors (Stu-

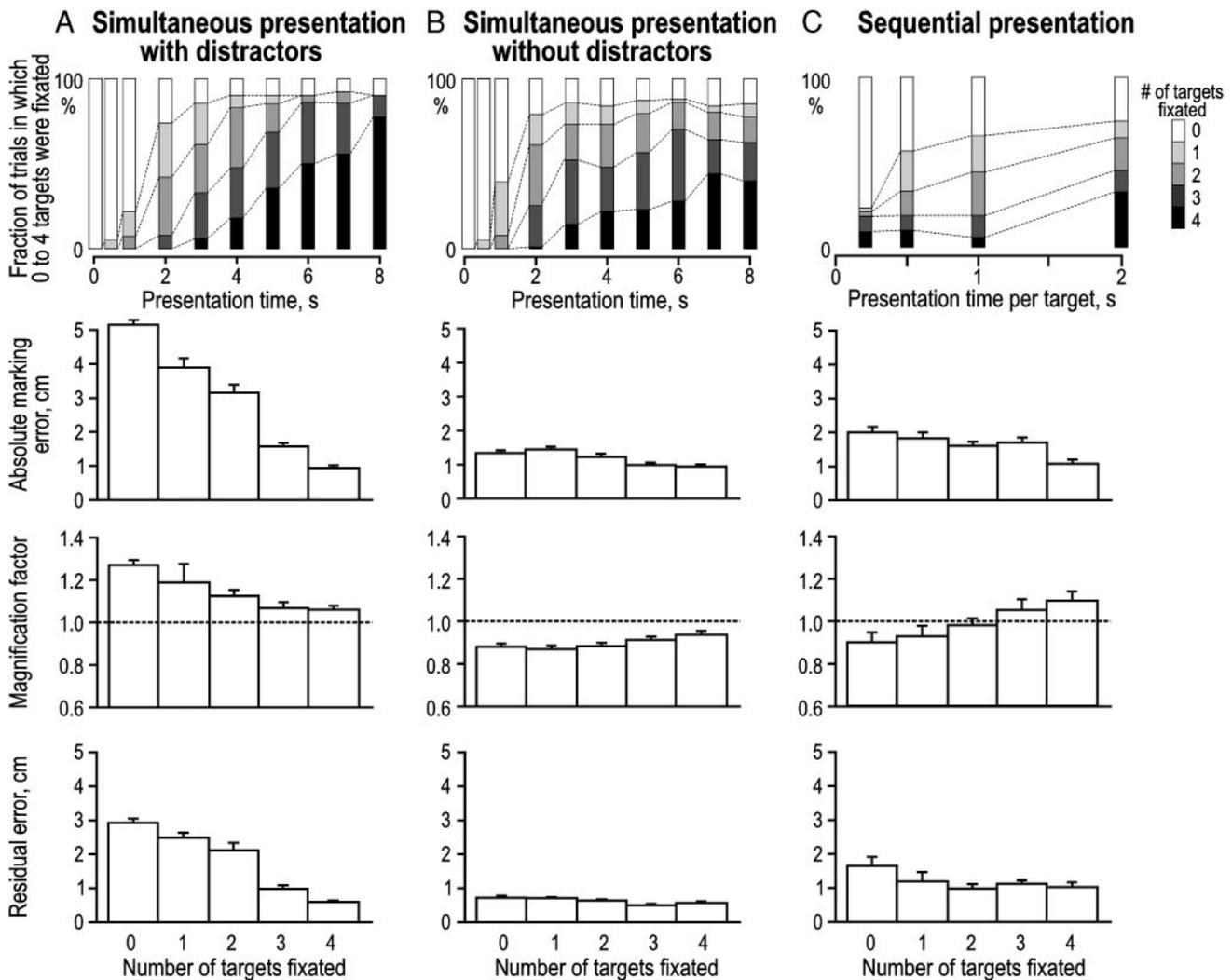


FIG. 4. Gaze behavior during encoding related to manual performance for simultaneous presentation with (A) and without (B) distractors and sequential presentation (C). *Top*: the fractions of trials in which 0, 1, 2, 3, and 4 targets were fixated during the encoding as a function of time of presentation. Each of the partitions in a column represent the fraction of trials in which a given number of targets were fixated; for key, see the inset linked to the C, *top*. *Top to bottom*: absolute marking error; magnification factor, and residual error as a function of the number of targets fixated during the encoding. Column height gives mean values and the error bar represents the standard error of mean for individual subjects.

dent's *t*-test:  $P = 0.78$ ). The magnification factor also decreased with the number of fixated targets [ $F(4,36) = 3.46$ ,  $P < 0.02$ ]. However, even when all four targets were fixated, it remained reliably greater than unity (Student's *t*-test:  $P < 0.03$ ). The residual error decreased reliably with the number of fixated targets [ $F(4,36) = 35.68$ ,  $P < 0.001$ ].

#### Manual performance after sequential target presentation without distractors

Our results indicate that with distractors, the subjects encoded the targets locations serially by successive gaze fixations whereas without distractors the subject could effectively encode the targets in parallel largely using parafoveal and peripheral vision. In this section, we analyze the manual performance when the targets were presented without distractors but sequentially instead of simultaneously. One interest here is whether the encoding of target locations depended on serial gaze fixations or whether they could effectively be encoded

largely by peripheral vision as during simultaneous presentation without distractors.

During sequential target presentation, the subjects rarely fixated any of the four targets at the shortest target presentation duration (0.2 s per target; Fig. 4C, *top*). With longer presentation times (0.5, 1, and 2 s per target), the subjects tended to fixate the targets following their appearance, although they rarely fixated all targets. Even at the longest presentation time (2 s per target), the subjects fixated all four targets in only 32.5% of the trials, while they fixated two or three targets in 19.0 and 12.3% of the trials. Thus although the targets were serially presented, subjects tended to encode their locations largely based on peripheral vision.

In contrast to simultaneous presentation without distractors, the absolute marking error was influenced by the number of targets fixated during the encoding period [ $F(4,36) = 10.58$ ,  $P < 0.001$ ; Fig. 3C, *2nd panel*). The absolute marking error increased as fewer targets were fixated, as during simultaneous presentation with distractors. However, even in trials where

none or only one target was fixated, the marking error was clearly smaller than when the corresponding number of targets was fixated during simultaneous presentation with distractors (cf. Fig. 4A and C). Target fixation also influenced the magnification factor [ $F(4,36) = 13.23, P < 0.001$ ], which increased with the number of targets fixated (Fig. 4C, 3rd panel). For trials in which the subjects did not fixate any of the targets, the magnification factor (Student's  $t$ -test:  $P = 0.90$ ) was similar to that with simultaneous presentation without distractors (Student's  $t$ -test:  $P = 0.90$ ). However, when they fixated three or more targets, the magnification factor was larger than that for simultaneous presentation without distractors (Student's  $t$ -test:  $P < 0.001$ ). Furthermore, for trials in which all four targets were fixated, the magnification factors during sequential presentation and simultaneous presentation with distractors were not reliably different (Student's  $t$ -test:  $P = 0.2$ ).

In sum, these results indicate that the marking of targets presented sequentially without distractors could benefit from serial target fixations in contrast to the parallel encoding of targets presented simultaneously without distractors. However, in the sequential target condition, the marking error only decreased modestly with the number of targets fixated. This may explain why the subjects often did not attempt to fixate the targets.

#### Transfer of sequential information between encoding and recall

We have demonstrated that gaze fixations of the targets improve the manual performance if the targets are serially encoded. Under such conditions, gaze fixations seem to play an important role for providing spatial information used to guide subsequent manual action. This raises the issue whether the sequential structure of the gaze program used to encode target locations transfer to the sequential structure of manual program used for recall.

For each trial, we compared the sequence with which the targets were fixated during the encoding period with the sequence by which the corresponding targets were marked. We confined the analysis to trials in which the subjects fixated all the targets in the series with sequential target presentation (12.4% of the trials) and simultaneous target presentation with distractors (11.4% of the trials). For trials in which the targets were presented simultaneously for longer periods, the subjects could refixate a target. Thus we also examined the relation between the sequence of targets marked during recall and the sequence of the last four targets fixated during encoding. However, this analysis yielded no evidence of sequence transfer and we therefore only report data pertaining to the first fixations observed during encoding.

For simultaneous presentation with distractors (Fig. 5A), the frequency distribution of the 24 possible combinations of encoding-recall sequences was significantly different from a uniform distribution ( $\chi^2$  test:  $P < 0.001$ ). Thus the sequence of marking was related to the sequence in which the targets were fixated. The most common marking sequence (29.1%) was the one in which the subjects marked the targets in the same sequence as the targets were visually encountered (i.e., 1234); the frequency of occurrence of this marking sequence was well above chance level (Fig. 5A). The reverse sequence, 4321 (11.8%) was the second most common but its occurrence did not exceed chance level. As with

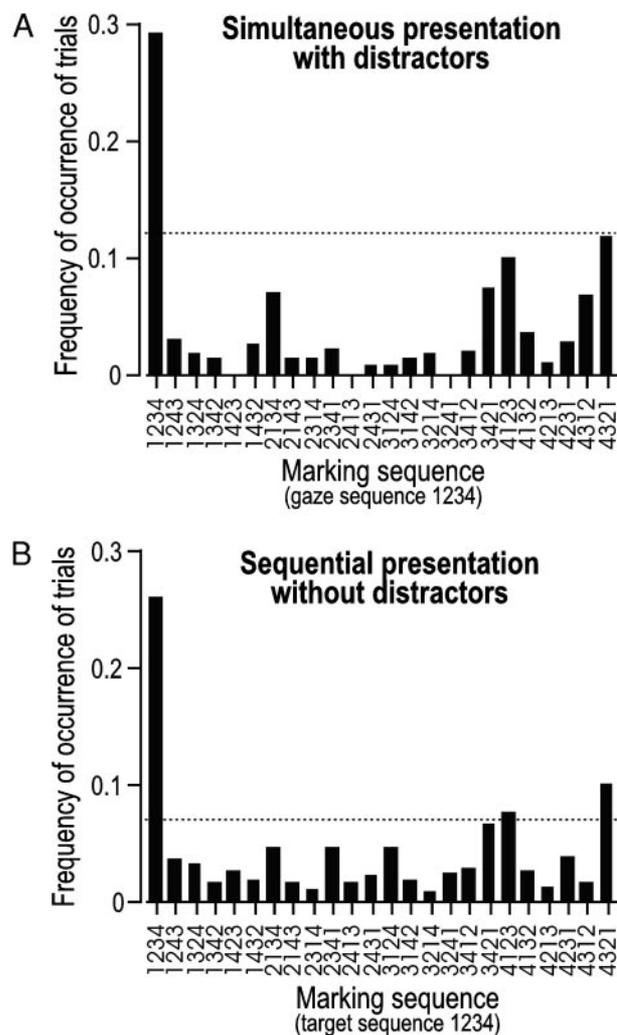


FIG. 5. Transfer of order information between gaze sequence and marking sequence with simultaneous presentation with distractors (A) and sequential presentation (B). The abscissa gives the 24 possible sequences of target marking with reference to the sequence by which the targets were initially fixated; 1, 2, 3, 4 denote the 1st, 2nd, 3rd, and 4th fixated target. The ordinate gives the relative frequency of occurrence of each marking sequence; only trials in which all 4 targets were fixated were analyzed. ---, the upper 95% confidence limit of frequency for the probability of occurrence of each of the 24 marking sequences if they would have occurred randomly, adjusted for multiple comparisons. A: data pooled across the 10 subjects who participated in the experiment for simultaneous presentation with distractors. B: data pooled across the 6 subjects who participated in the experiment for sequential presentation.

simultaneous target presentation, for the sequential presentation the frequency of occurrence of each of the 24 possible encoding-recall sequences was not distributed uniformly ( $\chi^2$  test:  $P < 0.001$ ) (Fig. 5B). The subjects marked the remembered locations of targets in the same sequence in which they were presented (sequence 1234) in 26.0% of the trials. Furthermore, in 10% of the trials, the subjects marked the targets in the reverse sequence (sequence 4321). The sequence 4123 (7.5%) also occurred at frequencies above chance, whereas the remaining 21 marking sequences all occurred below chance level. The distribution of marking sequences did not differ significantly from that obtained for the simultaneous presentation ( $\chi^2$  test:  $P = 0.32$ ; cf. Fig. 5, A and B). Thus the sequence in which the targets were encountered during the presentation period influenced to some degree the

sequence in which they were recalled during the marking. For simultaneous presentation without distractors, there was no clear relation between the sequence of target marking during recall and the gaze sequence during presentation (data not shown). That is, none of the 24 marking sequences with respect to target order occurred with a probability that was reliably different from chance ( $\chi^2$  test:  $P > 0.2$ ).

#### *Interval between target presentation and recall*

For six subjects, we varied the delay between target offset and manual response between 0.5 and 8.0 s, while setting the target presentation time at a constant value (8 s). The length of the delay did not significantly affect manual performance in any experimental condition, i.e., simultaneous target presentation with and without distractors. This indicates that the memory involved with this task was stable over the delay periods investigated (0.5–8 s). During the delay, we did not observe eye movements that seemed to serve as a “rehearsal” to the future marking movements. Furthermore, the intervals between consecutive markings were around 0.5 s regardless of the delay. There were no obvious influences of the time and mode of target presentation and the number of targets fixated on the pace of target markings.

#### DISCUSSION

With targets presented simultaneously without distractors, at the shortest presentation time (0.25 s), subjects did not fixate any target. Nevertheless, they encoded the target locations accurately at a level comparable to that for targets presented for 8 s where some three targets were typically fixated. This effective and quick marking of target location indicate that subjects encoded the targets in parallel largely using peripheral vision. On the other hand, when targets were presented simultaneously with distractors, the subjects had to fixate the individual targets with central vision to get their locations accurately, and thus the targets were encoded serially. The preferred gaze behavior matched these two modes of encoding. With distractors, subjects fixated nearly all targets when the presentation time was sufficiently long, whereas without distractors, subjects usually chose not to fixate all targets even when they had enough time to do so. Despite these differences in encoding strategy, the absolute marking error was similar for targets presented simultaneously without distractors and with distractors provided all targets were fixated in the latter case.

These two encoding strategies can be related to two previously proposed distinct visual search strategies for detection of targets among distractors in a visual scene (Treisman and Gormican 1988; Treisman and Souther 1985). In displays where the targets seemed to “pop out” distinctly from distractors, the targets are detected within a short time period that is independent of the number of distractors, suggesting parallel scanning of the scene. In contrast, in displays where the targets differ from the distractors only in specific details, as in the present study, the time required for visual search increases with the number of distractors. This suggests a serial scanning process in which items are evaluated in central vision. Note that in these studies by Treisman and colleagues, subjects reported the presence or absence of a target by pressing a button. In contrast, in the current study, subjects were required to recall the encoded target locations. Furthermore, our data

provide direct evidence for parallel and serial encoding strategies based on measurements of gaze behavior.

We observed different patterns of magnification errors across the three conditions we examined. Shrinkage in the visuomotor map was always observed when targets were presented simultaneously without distractors regardless of whether or not some or all of the targets were fixated. In contrast, shrinkage was not observed when distractors were present. When targets were presented sequentially (without distractors), shrinkage occurred when none of the targets were fixated, but the degree of shrinkage decreased with the number of fixated targets and, when all four targets were fixated, no shrinkage was observed. These results can be accounted for by two assumptions that the remembered location of any target encoded in peripheral vision is biased toward the fixation point and that this bias holds even for targets that have been previously fixated during the target presentation period. When targets were presented simultaneously among distractors, they could not be detected in peripheral vision (i.e., central vision was required to encode their locations), and thus no shrinkage was observed. When targets were presented sequentially and encoded in peripheral vision (i.e., not fixated), shrinkage occurred. However, when subjects fixated these targets, shrinkage disappeared. When targets were presented simultaneously without distractors, shrinkage was observed even when subjects fixated each target in turn. In this case (unlike the sequential target condition), when a given target was fixated, the other three remained in peripheral vision. If the representation of these peripheral targets is updated following each saccade and assuming that their remembered locations are biased toward the current fixation point, then shrinkage will be expected.

There is ample evidence that visual stimuli are transformed from retinal coordinates into motor coordinates by dynamically updating their spatial representations in conjunction with voluntary eye (or hand) movements (for a review, see Colby and Goldberg 1999). Neurons in the intermediate layers of the superior colliculus (Mays and Sparks 1980), the frontal eye field (Goldberg and Bruce 1990), and the lateral intraparietal area (LIP) (Barash et al. 1991a,b; Duhamel et al. 1992; Goldberg and Bruce 1990) exhibit signals that specify vanished saccade targets in coordinates of a new fixation made after target extinction. Similar signals are observed in the anterior intraparietal (AIP) area with respect to changes in hand position (Jeannerod et al. 1995; Rizzolatti et al. 1988, 1990; Sakata and Taira 1994). Neurons in LIP are also modulated by attention such that the neural response to a given stimulus is strongly influenced by its salience in a given task (Gottlieb et al. 1998; Kusunoki et al. 2000; Lynch et al. 1977; Mountcastle et al. 1981). Such dynamic updating of motorically represented target locations is presumably involved in the memory-guided marking tasks we have examined. For example, recent evidence suggests that the frontal eye field maintains a representation of the visual world that can last for several minutes—well within the delay periods we examined—and that is not dependent on continuous visual stimulation (Umeno and Goldberg 1997, 2001). When our subjects fixated all targets in the sequential presentation task or the simultaneous presentation with distractors task, the locations of previously fixated—and now undetectable—targets are presumably updated by such a mechanism. However, our results suggest that when previously fixated targets remain in peripheral vision, a different process

may be involved. Specifically, the location of such targets may be re-encoded based on peripheral vision.

A number of studies have observed that subjects underestimate the eccentricity of targets presented in peripheral vision during steady fixation (Mateeff and Gourevich 1983; Mitrani and Dimitrov 1982; Osaka 1977; Rauk and Luuk 1978). Moreover, a similar underestimation is observed when subjects are asked to shift their gaze to briefly presented eccentric targets (Cai et al. 1997; Honda 1993, 1995; Matin 1972; Ross et al. 1997). Our finding that shrinkage occurs when targets are encoded via peripheral vision is consistent with these results. Sheth and Shimojo (2001) found that the extent of underestimation or compression increases with the delay between target presentation and the response to indicate its position. However, we did not observe a change in the level of shrinkage across delay times.

#### *Transfer of sequential information between encoding and recall*

Noton and Stark (1971a,b) argued that memory for a scene not only contains spatial features but also information about the gaze sequence and suggested that the sequence of fixations in initial viewing of the scene and later recognition should be similar. However, we found little evidence that the same applies to manipulatory tasks in which the recall is expressed in sequential manual actions directed towards multiple targets. For the simultaneous presentation with distractors where subjects were free to select the sequence of gaze and for the sequential presentation, there was only a weak tendency for the marking sequence to be in the same (or reverse) order as the targets were fixated during the encoding period. This suggests that when visually encoding future targets for manual sequential actions, the central nervous system build a representation of their locations that is largely independent of the gaze sequence and the hand action sequence. That is, the gaze sequences fixating objects to be manipulated in the future seem neither to constrain the order of forthcoming manual sequences nor to be influenced by an action sequence preferred by the neural apparatus that plan and control hand actions.

We thank Dr. G. Westling and A. Bäckström for technical assistance.

This study was supported by the Swedish Medical Research Council (Project 08667), the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, the Fifth Framework Program of European Union (project: QLG3-CT-1999-00448), Sankyo Foundation of Life Science, and the Canadian Institutes of Health Research.

#### REFERENCES

- ABRAMS RA, MEYER DE, AND KORNBUM S. Eye-hand coordination: oculomotor control in rapid aimed limb movements. *J Exp Psychol* 16: 248–267, 1990.
- BALLARD DH, HAYHOE MM, LI F, AND WHITEHEAD SD. Hand-eye coordination during sequential tasks. *Philos Trans R Soc Lond B Biol Sci* 337: 331–338, 1992.
- BARASH S, BRACEWELL RM, FOGASSI L, GNADT JW, AND ANDERSEN RA. Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J Neurophysiol* 66: 1095–1108, 1991a.
- BARASH S, BRACEWELL RM, FOGASSI L, GNADT JW, AND ANDERSEN RA. Saccade-related activity in the lateral intraparietal area. II. Spatial properties. *J Neurophysiol* 66: 1109–1124, 1991b.
- BIGUER B, PRABLANC C, AND JEANNEROD M. The contribution of coordinated eye and head movements in hand pointing accuracy. *Exp Brain Res* 55: 462–469, 1984.
- BINSTED G AND ELLIOT D. Ocular perturbations and retinal/extraretinal information: the coordination of saccadic and manual movements. *Exp Brain Res* 127: 193–206, 1999.
- BOCK O. Contribution of retinal versus extraretinal signals towards visual localization in goal-directed movements. *Exp Brain Res* 64: 476–482, 1986.
- CAI RH, POUGET E, SCHLAG-REY M, AND SCHLAG J. Perceived geometrical relationships affected by eye movement signals. *Nature* 387: 601–604, 1997.
- COLBY CL AND GOLDBERG ME. Space and attention in parietal cortex. *Annu Rev Neurosci* 22: 319–349, 1999.
- DUHAMEL JR, COLBY CL, AND GOLDBERG ME. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255: 90–92, 1992.
- GOLDBERG ME AND BRUCE CJ. Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. *J Neurophysiol* 64: 489–508, 1990.
- GOTTLIEB JP, KUSUNOKI M, AND GOLDBERG ME. The representation of visual salience in monkey parietal cortex. *Nature* 391: 481–484, 1998.
- GOOLSBY TW. Eye movement in music reading: effects of reading ability, notational complexity, and encounters. *Music Percept* 12: 71–96, 1994.
- HODGSON TL, BAJWA A, OWEN AO, AND KENNARD C. The strategic control of gaze direction in the Tower of London task. *J Cognit Neurosci* 12: 894–907, 2000.
- HONDA H. Saccade-contingent displacement of the apparent position of visual stimuli flashed on a dimly illuminated structured background. *Vision Res* 33: 709–716, 1993.
- HONDA H. Visual mislocalization produced by a rapid image displacement on the retina: examination by means of dichoptic presentation of a target and its background scene. *Vision Res* 35: 3021–3028, 1995.
- INHOFF AW AND WANG J. Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *J Exp Psychol Hum Percept Perform* 18: 437–448, 1992.
- JOHANSSON RS, WESTLING G, BÄCKSTRÖM A, AND FLANAGAN R. Eye-hand coordination in object manipulation. *J Neurosci* 21: 6917–6932, 2001.
- JEANNEROD M, ARBIB MA, RIZZOLATTI G, AND SAKATA H. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci* 18: 314–320, 1995.
- KINSLER V AND CARPENTER RH. Saccade eye movements while reading music. *Vision Res* 35: 1447–1458, 1995.
- KOWLER E. Eye Movements. In: *Visual Cognition*, edited by Kosslyn SM, and Osherson DN. Cambridge, MA: MIT Press, 1995, vol. 2, p. 215–266.
- KUSUNOKI M, GOTTLIEB J, AND GOLDBERG ME. The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. *Vision Res* 40: 1459–1468, 2000.
- LAND MF. Predictable eye-hand coordination during driving. *Nature* 359: 318–320, 1992.
- LAND MF AND FURNEAUX S. The knowledge base of the oculomotor system. *Philos Trans R Soc Lond B Biol Sci* 352: 1241–1239, 1997.
- LAND M AND HORWOOD J. Which parts of the road guide steering? *Nature* 377: 339–340, 1995.
- LAND MF AND LEE DN. Where we look when we steer. *Nature* 369: 742–744, 1994.
- LAND MF AND McLOED P. From eye movements to actions: how batsmen hit the ball. *Nat Neurosci* 3: 1340–1345, 2000.
- LAND M, MENNIE N, AND RUSTED J. The roles of vision and eye movements in the control of activities in daily living. *Perception* 28: 1311–1328, 1999.
- LYNCH JC, MOUNTCASTLE VB, TALBOT WH, AND YIN TCT. Parietal lobe mechanisms for directed visual attention. *J Neurophysiol* 40: 362–389, 1977.
- MATEEFF S AND GOUREVICH A. Peripheral vision and perceived visual direction. *Biol Cybern* 49: 111–118, 1983.
- MATIN L, JAMESON D, AND HURVICH L. In: *Handbook of Sensory Physiology*. Berlin, Germany: Springer, 1972, vol. 7:4, p. 144–153, p. 307–332.
- MAYS LE, SPARKS DL. Dissociation of visual and saccade-related responses in superior colliculus neurons. *J Neurophysiol* 43: 207–232, 1980.
- MITRANI L AND DIMITROV G. Retinal and visual localization during pursuit eye movements. *Vision Res* 22: 1047–1051, 1982.
- MOUNTCASTLE VB, ANDERSEN RA, AND MOTTER BC. The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J Neurosci* 7: 1218–1235, 1981.
- NEGGERS SFW AND BEKKERING H. Ocular gaze is anchored to the target of an ongoing pointing movement. *J Neurophysiol* 83: 639–651, 2000.
- NOTON D AND STARK L. Scan paths in eye movements during pattern perception. *Science* 171: 308–311, 1971a.

- NOTON D AND STARK L. Scan paths in saccadic eye while viewing and recognizing patterns. *Vision Res* 11: 929–942, 1971b.
- OSAKA N. Effect of refraction on peripheral locus of a target in peripheral visual field. *J Psychol* 95: 59–62, 1977.
- PATLA AE AND VICKERS JN. Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport* 8: 3661–3665, 1997.
- PRABLANC C, PÉLISSON D, AND GOODALE MA. Visual control of reaching movements without vision of the limb. I. Role of retinal feedback of target position in guiding the hand. *Exp Brain Res* 62: 293–302, 1986.
- PÉLISSON D, PRABLANC C, GOODALE MA, AND JEANNEROD M. Visual control of reaching movements without vision of the limb. II. Evidence of fast unconscious processes correcting the trajectory of the hand to the final position of a double-step stimulus. *Exp Brain Res* 62: 303–311, 1986.
- PYLYSHYN ZW. Situating vision in the world. *Trends Cognit Sci* 4: 197–207, 2000.
- RAUK M AND LUUK A. Perceived visual direction of the brief test-flashes on the horizontal scale. *Acta Comment Univer Tartuensis* 474: 900–1000, 1978.
- RIZZOLATTI G, CAMARDA R, FOGASSI L, GENTILUCCI M, LUPPINO G, AND MATELLI M. Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp Brain Res* 71: 491–507, 1988.
- RIZZOLATTI G, GENTILUCCI M, CAMARDA RM, GALLESE V, LUPPINO G, MATELLI M, AND FOGASSI L. Neurons related to reaching-grasping arm movements in the rostral part of area 6 (area 6a beta). *Exp Brain Res* 82: 337–350, 1990.
- ROSS SG, MORRONE MC, AND BURR DC. Compression of visual space before saccades. *Nature* 386: 598–601, 1997.
- SAKATA H AND TAIRA M. Parietal control of hand action. *Curr Opin Neurobiol* 4: 847–856, 1994.
- SHETH BR AND SHIMOJO S. Compression of space in visual memory. *Vision Res* 41: 329–341, 2001.
- SMEETS JB, HAYHOE MM, AND BALLARD DH. Goal-directed arm movements change eye-hand coordination. *Exp Brain Res* 109: 434–440, 1996.
- TREISMAN A. Solutions to the binding problem: progress through controversy and convergence. *Neuron* 24: 105–110, 1999.
- TREISMAN A AND GORMICAN S. Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev* 95: 15–48, 1988.
- TREISMAN A AND SOUTHER J. Search asymmetry: a diagnostic for preattentive processing of separable features. *J Exp Psychol Gen* 114: 285–310, 1985.
- UMENO MM AND GOLDBERG ME. Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J Neurophysiol* 78: 1373–1383, 1997.
- UMENO MM AND GOLDBERG ME. Spatial processing in the monkey frontal eye field. II. Memory responses. *J Neurophysiol* 86: 2344–2352, 2001.
- VICKERS JN. Gaze control in putting. *Perception* 21: 117–132, 1992.
- VICKERS JN. Visual control when aiming at a far target. *J Exp Psychol Hum Perform* 22: 342–354, 1996.