Spatial Transformations for Eye–Hand Coordination

J. D. Crawford,1,2 W. P. Medendorp,3 and J. J. Marotta1,2

1Canadian Institutes of Health Research Group for Action and Perception, 2York Centre for Vision Research, Departments of Psychology, Biology, and Kinesiology and Health Sciences, York University, Toronto, Ontario M3J 1P3, Canada; and 3Nijmegen Institute for Cognition and Information and FC Donders Centre for Cognitive Neuroimaging, University of Nijmegen, NL 6500 HE, Nijmegen, The Netherlands

Submitted 4 February 2004; accepted in final form 20 February 2004

Crawford, J. D., W. P. Medendorp, and J. J. Marotta. Spatial transformations for eye–hand coordination. J Neurophysiol 92: 10–19, 2004; 10.1152/jn.00117.2004. Eye–hand coordination is complex because it involves the visual guidance of both the eyes and hands, while simultaneously using eye movements to optimize vision. Since only hand motion directly affects the external world, eye movements are the slave in this system. This eye–hand visuomotor system incorporates closed-loop visual feedback but here we focus on early feedforward mechanisms that allow primates to make spatially accurate reaches. First, we consider how the parietal cortex might store and update gaze-centered representations of reach targets during a sequence of gaze shifts and fixations. Recent evidence suggests that such representations might be compared with hand position signals within this early gaze-centered frame. However, the resulting motor error commands cannot be treated independently of their frame of origin or the frame of their destined motor command. Behavioral experiments show that the brain deals with the nonlinear aspects of such reference frame transformations, and incorporates internal models of the complex linkage geometry of the eye–head–shoulder system. These transformations are modeled as a series of vector displacement commands, rotated by eye and head orientation, and implemented between parietal and frontal cortex through efficient parallel neuronal architectures. Finally, we consider how this reach system might interact with the visually guided grasp system through both parallel and coordinated neural algorithms.

INTRODUCTION

Eye–hand coordination is central to so many human activities—tool use, eating, sports, and work, to name a few—as to be a defining characteristic of typical human life. Conversely, its disruption following stroke, disease, injury, and developmental disorders leads to a considerable degeneration in productivity and quality of life. Normal eye–hand coordination involves the synergistic function of several sensorimotor systems, including the visual system, vestibular system, proprioception, and the eye, head, and arm control systems, plus aspects of cognition-like attention and memory. This makes understanding the neural underpinnings of eye–hand coordination rather daunting, even if we consider it to be only the sum of its parts. Eye–hand coordination is still more than this, however; it evokes combinatorial problems that do not arise when we study the individual component systems in isolation. In the end though, the purpose of the eye–hand coordination “system” is straightforward: the use of vision to guide movements of the hand (reaching, grasping, and manipulation). Remembering this fundamental fact is our best tool in understanding the function of the whole system.

The focus of this review will be on the spatial aspects of how vision is transformed into hand motion within the context of a system in which the eyes (and head) are also moving constantly to optimize vision-for-action. Specifically, we will consider how the brain deals with the geometric problems of transforming a stimulus coded at the level of the retina into a motor code useful for controlling reaching and grasping motions. First, by way of background, we will briefly review the behavioral aspects of eye–hand coordination and the relative roles of visual feedback and feedforward mechanisms in arm control.

Behavioral aspects of eye–hand coordination

The spatiotemporal relationships between eye and hand movements in natural behavior are complex (Furneaux and Land 1999; Hayhoe et al. 2003; Herst et al. 2001; Land and McLeod 2000; Peltz et al. 2001), but again are probably best understood in terms of optimizing vision for the guidance of hand motion (Johansson et al. 2001; Regan and Gray 2001; Steinman et al. 2003). The temporal coupling of eye and hand movements varies in a task-dependent manner, presumably to optimize the useful flow of visual information for a particular task (Fisk and Goodale 1985; Land and Hayhoe 2001; Rossetti et al. 1993; Sailer et al. 2000). Reaching toward and manipulating objects is degraded when gaze is deliberately deviated from its normal sequence of target fixations (Henriques and Crawford 2000; Henriques et al. 2003; Terao et al. 2002), and in at least some cases gaze seems locked on the target until it is reached by the hand, independent of visual feedback (Neggers and Bekkering 2000, 2001).

Such gaze fixation strategies are useful because they place the visual target on the part of the retina (the fovea) with the most densely packed sensory apparatus, while temporarily removing the added burden of spatial updating for gaze shifts (see Gaze-centered representations and spatial updating section below). Moreover, fixating gaze at particularly task-relevant points in a coordinated sequence allows for periods in which the brain can calculate the geometric relationships between the external world (through vision) and the internal world through proprioception (Johansson et al. 2001).

Gaze and arm movements sometimes appear to be guided by a common drive signal, for example, being influenced in similar ways by visual illusions and in tracking strategies (Engel et
al. 2000; Soechting et al. 2001). Likewise, movements of the eyes and arm influence each other’s kinematic profiles (Epelboim et al. 1997; Fisk and Goodale 1985; Snyder et al. 2002; Van Donkelaar 1998), presumably revealing mutual triggering or facilitating mechanisms between oculomotor and prehensile circuits located within specific brain regions (Carlton et al. 2002; Miall et al. 2001; Van Donkelaar et al. 2000). Yet at other times, eye and arm movements are naturally decoupled (Fischer et al. 2003; Henriques et al. 2003; Steinman et al. 2003), at least in healthy individuals (Carey et al. 1997).

Again, these rules and their exceptions likely emerge from the task-dependent use of vision to guide eye and arm movements, while simultaneously programming eye movements to optimize vision (Bekkering and Niggens 2002; Steinman et al. 2003). Next, we turn to considering how the brain implements just one aspect of this recursive sequence, the visual guidance of arm movement.

Feedback versus feedforward

In robot control systems, the complexity of control is significantly reduced by using visual feedback to “visually servo” the effector, essentially driving it to the point where visual error is reduced to zero (e.g., Kragic et al. 2002). This works in robotics because sensory feedback is limited only by the speed of electrical flow and computer processing time. In the real primate brain though, the speed of neural conduction and processing time is such that a rapid saccadic eye movement would be finished, or a fast arm movement would be way off track before it was accurately updated by a new visual signal (e.g., Robinson 1981). So, the eye–hand coordination system must either rely completely on this slow sensory feedback and make very slow movements (maybe the brain of the South American tree sloth has gone for this option), or it must take another route: the use of internal models of the physical system and external world that, based on initial sensory conditions, can operate with some subsequent independence.

This is not to say that visual feedback is not used to guide reaching and grasping movements. Visual feedback alters reaching kinematics (Connoly and Goodale 1999) even without conscious perception (Goodale et al. 1986; Prablanc and Martin 1992) and we must rely on such feedback when engaging in new behaviors or when we encounter unexpected conditions (Baddeley et al. 2003; Flanagan et al. 2003; Johansson et al. 2001; Rossetti et al. 1993). There is evidence that posterior parietal cortex (PPC) also helps to incorporate visual feedback into on-going arm movements (Desmurget et al. 1999; Pisella et al. 2000). Even well-practiced movements require visual feedback for optimal performance (Proteau and Carnahan 2001). However, feedforward transformations are essential for the basic aspects of common, overlearned behaviors (Ariff et al. 2002; Flanagan et al. 2001, 2003), allowing intermittent visual fixations to accurately guide a rapid, continuous sequence of coordinated eye and arm movements. Subsequent sections of this review will deal with the level of sophistication that is attained in these feedforward internal models.

Even with the use of such feedforward internal models, the internal structure of the brain is massively recursive. It has correctly been stated that the cortical structures involved in the visuomotor transformations for arm movement are nested within loops, making them more like an interdependent system than a set of discrete transformations (e.g., Caminiti et al. 1998). However, if we hope to understand what this system does, we need to start by dividing the transformations into conceptual steps and then attempt to divine how these steps might be implemented. To save time the brain presumably implements sequential computations using the shortest possible paths. Coupled with this, the primate brain appears to be organized into certain modular computational units (e.g., Andersen and Buneo 2002; Wise et al. 1997). Thus there is hope that we can identify some of the feedforward transformations for eye–hand coordination.

Because the “late” aspects of these transformations—certain inverse kinematic transformations, calculations of muscle dynamics, short-loop proprioceptive feedback reflexes—are linked to the ubiquitous control of limb movement rather than eye–hand coordination per se, our focus here will mainly be on the “early aspects” (Flanders et al. 1992), the incorporation of visual information into the motor plan, and how this compensates for movements of the eyes and head.

Early visual representations for arm movement

A central concept in the use of internal models for eye–hand coordination is that of reference-frame transformations (Flanders et al. 1992). It is often stated that vision-for-reaching is initially coded in a gaze (or eye)-centered frame and that this must eventually be transformed into a hand-centered frame. However the latter part of this statement is incorrect. The final stage of eye–hand coordination is muscular contraction. Although hand position is the controlled variable, the main muscles that control it have their stable insertion points in the upper arm and shoulder, so these are the final frames of reference for eye–arm coordination.

Because eye–hand control begins from a retinal frame, the classic problem arises that every time the eyes (and head) shift(s) gaze, they interrupt vision and disrupt the spatial relationship between the sensory apparatus and the external world (Hallett and Lightstone 1976). The system could wait until the gaze shift is finished to update its visual information (O’Regan and Noe 2001), but gaze shifts often take the original target of interest from the high-resolution fovea to the less-sensitive peripheral retina, and sometimes even out of the visual range. Perhaps more important, reliance on external feedback would introduce redundant visual computations and long lags in processing time (saccade time + visual reprocessing time ≈ 250 ms) rendering eye–hand coordination inefficient, and visual guidance of arm movements during a rapid sequence of saccades nearly impossible. To avoid this, representations that are important for future actions must be stored, either in a form that is independent of eye movement, or internally updated to compensate for the eye movement (Duhamel et al. 1992).

Gaze-centered representations and spatial updating

It is thought that the eye–hand coordination system constructs both egocentric and allocentric representations of visual space, depending on various factors including the available sensory information, the task constraints, the visual background, memory interval, and the cognitive context (Battaglia-Mayer et al. 2003; Hayhoe et al. 2003; Hu and Goodale 2000).
In an otherwise neutral space, however, a simple viewer-centered coordinate system appears to be used for the early planning of reaching and pointing\(^1\) targets (McIntyre et al. 1997; Vetter et al. 1999). Until recently, however, it was unclear how the system stored these early motor representations, moment to moment, during eye movements.

Henriques et al. (1998) addressed this question in an open-loop pointing task, where subjects had to foveate a briefly flashed target, then deviate their eyes, and then to point toward the remembered location in complete darkness (Fig. 1). Henriques et al. (1998) sought to determine whether such pointing responses were affected by an intervening eye displacement by comparing them to pointing to remembered foveal targets (control condition) or to retinally peripheral targets (static condition). The idea of the test was that if subjects were pointing using a nonretinal representation, intervening eye movements would have no effect, and responses as in the control condition would be expected. However, if subjects were pointing based on an updated gaze-centered representation, pointing behavior would echo pointing to peripheral targets. The results clearly supported the latter. As shown by Fig. 1 (right column), this was found for targets independent of their distance from the subject (Medendorp and Crawford 2002). Recently, such gaze-centered updating was also reported for pointing to auditory and proprioceptive targets (Pouget et al. 2002) and for pointing to the center of wide-field expanding motion patterns (Poljac and Van den Berg 2003).

Subsequent neurophysiological studies are consistent with these psychophysical findings. Batista et al. (1999) demonstrated that the monkey parietal reach region (PRR)—an arm control center in the PPC—uses retinocentric receptive fields and a gaze-centered updating mechanism. This is consistent with visuospatial processing and movement planning in other, more saccade-related areas, including extrastriate visual areas (Nakamura and Colby 2002), the lateral intraparietal area (Duhamel et al. 1992), the frontal eye fields (Unemo and Goldberg 1992), and the superior colliculus (Walker et al. 1995).

Recently, a human analog of PRR has been identified (Connolly et al. 2003) and 2 functional magnetic resonance imaging (fMRI) studies have reported evidence for spatial updating in the human parietal cortex in conjunction with eye movements (Medendorp et al. 2003b; Merriam et al. 2003). Medendorp et al. took as their starting point a previously reported bilateral region in the human PPC that shows contralateral topography for memory-guided eye movements (Sereno et al. 2001). They showed further that this region, illustrated in Fig. 2, is also activated for arm movements. To demonstrate updating, they used event-related fMRI and showed that stored memory activity in this region for both eye and arm movements is dynamically remapped between the 2 hemispheres when eye movements cause target remapping relative to the gaze fixation point. This suggests that much of the previous physiological work done on monkey PPC also applies to the human.

\[^1\] For our purposes here, pointing movements can be equated to a reach without the grasp component, since humans do not point to distant targets along a direct line from the shoulder but rather as if reaching toward a retinal stimulus located at arm’s length (Flanders et al. 1992; Henriques and Crawford 2002).
us to a point often ignored in the motor control literature: we have two eyes, not one. How is binocular information synthesized for arm control?

This topic is contentious (Erkelens and Van Ee 2002), but the classic visual perception literature suggests that binocular information is fused and referenced to a virtual cyclopean eye located midway between the left and right eyes (e.g., Ono et al. 2002). One needs to be careful in extrapolating perceptual data to motor control because it is now thought that the visual brain uses separate analytic streams for perception and motor control (Goodale and Milner 1992). However, the idea of an egocentric perceptual reference point agrees with the motor-based finding that 3-D errors in visually guided reaching form 3-D ellipses whose long axes converge toward some point on the face

This technique relies on the assumption that pointing errors in depth, elevation, and azimuth are independent (no cross talk), that errors are greatest along one dimension (depth), and that the errors arising from a single point in the sequence of transformations. Violations of the latter assumption probably account for the different result in pointing without visual feedback, which produces ellipses converging on the shoulder (Flanders et al. 1992; Soechting et al. 1990), perhaps reflecting errors in planning the movement path more than errors representing target location.

2 This technique relies on the assumption that pointing errors in depth, elevation, and azimuth are independent (no cross talk), that errors are greatest along one dimension (depth), and that the errors arising from a single point in the sequence of transformations. Violations of the latter assumption probably account for the different result in pointing without visual feedback, which produces ellipses converging on the shoulder (Flanders et al. 1992; Soechting et al. 1990), perhaps reflecting errors in planning the movement path more than errors representing target location.
Actually, displacements are not frame-independent (Crawford and Guittion 1997). Displacement is independent of relative translocations of different frames, but this is not what the eye, head, and body do. Primarily, they rotate with respect to each other. This results in potentially huge differences in the displacements coded in these different frames (Fig. 3A–C).

We tested to see whether subjects accounted for these differences by having them point between horizontally displaced targets (Fig. 3D) flashed in the dark with the eyes fixated at different vertical elevations (Crawford et al. 2000). Using 3-D eye coil signals, we calculated the location of the targets in retinal coordinates, and used this as input to the direct transformation model (although it was not called such at that time) to generate quantitative predictions. In retinal coordinates, the horizontal target displacements also had vertical components (fanning outward as a nonlinear function of vertical eye orientation (Fig. 3E)), so the direct transformation model predicted a similar “fanning out” pattern of arm movement errors (Fig. 3F). One subject showed a tendency toward this pattern, but most subjects clearly incorporated the nonlinear, eye orientation-dependent transformation required for ideal behavior (Fig. 3G). Further, one needs to incorporate similar transformations for head orientation, or else arm movements would be entirely inaccurate (Klier et al. 2001).

Thus even after reach targets are updated for intervening motion of the eyes and head, whether they are compared with hand position in an early retinal frame or at some later stage, a second set of reference frame transformations is still required for accurate reach control (Henriques et al. 1998).

Accounting for eye–head–shoulder linkage geometry

The importance of an “egocentric reference location” (see Representing reach targets in 3-D space) for motor control becomes evident when one considers the linkage geometry of the eyes, head, and shoulder. If they all rotated about the same point (impossible) this geometry would be trivial. Because they do not, rotations of the head cause the eye (cyclopean or real) to translate with respect to the shoulder (Fig. 4.), considerably complicating the geometry of comparing signals in these 2 frames in either forward or reverse computations. For example, if the system relied only on angular eye orientation and head orientation to map visual target location onto a reach, and failed to compensate for the eye translation component, it would generate erroneous reach patterns at noncentral head positions (Fig. 4B). To solve this problem, the brain must estimate how the visual ego-center is translocated (in shoulder coordinates) during head rotations.

We tested to see whether the internal models for eye–hand coordination account for this geometry by having subjects point (Henriques et al. 2002) or reach (Henriques et al. 2003) toward briefly flashed targets at various distances in the dark, with the head at various horizontal angles (Fig. 4). Subjects were able to reach correctly as long as gaze was fixated on the

**FIG. 3.** Transformation of displacements from eye coordinates into space coordinates. A: imagine an eye at primary position (where the eye, head, and body are facing leftward in the graph). Consider 2 external reach targets, one straight ahead of the eye (■) and one to the eye’s left, coming out of the page (□). 3-D vector displacement between these 2 targets is shown by the dashed line (—). Now, treating the eye and the targets as one rigid body, rotate them either 90° upward (B), or 90° clockwise (CW) about the line of sight (C) (e.g., by a combined eye and head rotation that leave the body fixed in position). By definition the vector displacement (—) stays the same in eye coordinates, but in each case it is entirely different in body coordinates. In other words, hand motor error in eye coordinates cannot be used to estimate hand motor error in body coordinates without knowledge of eye and head orientation. D–G: experimental test of the situation shown in A–B, using real data modified from Crawford et al. (2000). D: spatial location of 5 horizontal target pairs at 5 vertical elevations, plotted in angular “eye coil” coordinates, viewed as though behind the subject. Task will be to fixate and point at the leftward member of each pair (head immobilized) and then point horizontally to the rightward member of each pair. E: rotated into eye coordinates (where 0, 0 = looking down the line of sight), the same 5 target displacements are now nonhorizontal, as a function of vertical eye orientation. F: when these “retinal errors” and initial arm positions are input to a direct transformation model, it predicts a “fanning out” pattern of pointing errors (gray wedges). G: actual arm trajectories to flashed rightward targets in the dark. Subjects did not make the errors predicted by the direct transformation model, but instead reached correctly to the correct arm positions (□), demonstrating that their visuomotor transformation incorporated knowledge of eye orientation.
motor behaviors; the real brain probably interchanges information between the early retinal maps and other more allocentric maps when required to interact with complex visual environments (Battaglia-Mayer et al. 2003; Hayhoe et al. 2003; Hu and Goodale 2003).

Figure 5 incorporates much of the discussion in the previous two sections. This model is similar to the scheme we had previously proposed (Crawford et al. 2003; Henriques et al. 1998), but incorporates the findings of Buneo et al. (2002). First of all, 3-D representations of target direction are stored and updated in a retinal frame. This corresponds to the first “representational stage” of our scheme. Second, these are compared with representations of hand location, transformed into retinal coordinates according to the scheme of Buneo et al. (2002) to compute a 3-D hand displacement in retinal coordinates. Importantly, however, our scheme still then requires a series of reference frame transformations of the motor displacement command from gaze coordinates into shoulder coordinates, by nonlinear comparisons with eye and head orientation. Again, this is necessary to reflect the actual 3-D geometry of the eye–head–shoulder system (Figs. 3 and 4).

Contrary to some suggestions (e.g., Hayhoe et al. 2003), models like this do not fail and cannot be disproved by the accurate performance of subjects in remembering multiple reach targets in complex natural environments. In theory, any number of targets can be represented in the original retinal frame, each to be converted “on demand” into an arm-movement command when required (Henriques et al. 1998). In fact, this has already been demonstrated in mathematical simulations (Medendorp et al. 2003a). In such models, it is only the overall series of transformations that provide a virtual representation of reach targets in external space. Having said this, though, we concede that gaze-centered visuomotor maps, by themselves, are probably not the most efficient means of storing complex information in real biological systems. Figure 5 likely represents the default transformation for simple visuo-

FIG. 4. Simplified linkage geometry of the eye, shoulder, and head. Lines of gaze (−−) from the 2 eyes converge on a target (*) within reach. The virtual “cyclopean eye” falls midway between the two eyes. Linkage between the right eye, head, shoulder, and arm are indicated by dashed lines (−−−) connecting the centers of rotation (○). A: accurate pointing with the head in a central orientation. B: rotating the head 40° left (while rotating gaze rightward toward the target) shifts the cyclopean eye slightly back and to the left (−−−). As a result, a reach based on angular information alone, without compensating for this shift, would overestimate the rightward location of the target (−−). Modified from data published in Henriques et al. (2003).

target; deviations of gaze apparently caused the algorithm to break down to some extent. This suggests that the internal models for eye–hand coordination achieve considerable geometric sophistication, at least within the range where they are likely to be well calibrated by use.

**Conceptual and physiological models for visually guided reach**

To illustrate the model, consider the following “task”: a subject looks at a briefly flashed target (●) with the arm at resting position (A). Then (B) the subject makes 1) an upward eye movement, followed by 2) a reaching or pointing movement toward the remembered target location (○). We hypothesize that the brain uses the following stages to do this. C: an early representational stage. Target location is stored in eye coordinates, such that this representation (○) must be counterrotated (updated) when the eye rotates. D: comparison stage. Updated target representation (○) is compared with an eye-centered representation of current hand location to generate “hand motor error” in eye coordinates (Buneo et al. 2002). E: visuomotor execution stage. “Hand motor error” signal is rotated by eye orientation and head orientation (or perhaps by gaze orientation) to put it into a body coordinate system appropriate for calculating the detailed inverse kinematics and dynamics of the movement. This last stage would also have to include internal models of the geometry illustrated in Fig. 4.

**FIG. 5.** Conceptual scheme for spatial transformations in eye–hand coordination. To illustrate the model, consider the following “task”: a subject looks at a briefly flashed target (●) with the arm at resting position (A). Then (B) the subject makes 1) an upward eye movement, followed by 2) a reaching or pointing movement toward the remembered target location (○). We hypothesize that the brain uses the following stages to do this. C: an early representational stage. Target location is stored in eye coordinates, such that this representation (○) must be counterrotated (updated) when the eye rotates. D: comparison stage. Updated target representation (○) is compared with an eye-centered representation of current hand location to generate “hand motor error” in eye coordinates (Buneo et al. 2002). E: visuomotor execution stage. “Hand motor error” signal is rotated by eye orientation and head orientation (or perhaps by gaze orientation) to put it into a body coordinate system appropriate for calculating the detailed inverse kinematics and dynamics of the movement. This last stage would also have to include internal models of the geometry illustrated in Fig. 4.
Models like this are useful in telling us what kind of signals are required for a transformation, but do not always tell us how they will be coded in the real brain. Clearly, the real brain does not use all of the explicit intermediate representations shown in Fig. 5, and whatever representations it does use are nested within feedback loops that link multiple brain regions (Caminiti et al. 1998). Based on this, one might despair that such transformations are implemented in no particularly recognizable order, or perhaps in one convergent tangle of signals (Battaglia-Mayer et al. 2000). However, Fig. 5 also incorporates the important fact that the 3-D reference frame transformations for visually guided movement are noncommutative; that is, they require nonlinear order-dependent calculations (Tweed et al. 1999). This transcends the capacity of simple summing junctions and places further constraints on the neural solutions. Probably the fastest way for a biological system to observe these constraints is through the use of feedforward parallel processing architectures (Smith and Crawford 2001a), leaving the longer, slower interregional feedback loops to play other roles, such as updating the system about its progress during the movement. So where then should one look in the brain for those intermediate forward transformations?

We have already seen that the PRR provides an eye-centered representation of the target direction that is updated during eye movements in both monkeys and humans (Batista et al. 1999; Medendorp et al. 2003a,b). The work of Buneo et al. (2002) provides a possible physiological substrate for the gradual transformation of hand position information into retinal coordinates and comparison with target information. What remains to be seen is how a gaze-centered estimate of desired hand motion would be transformed, into the shoulder- and arm-centered representations of hand motor error observed at the level of parietal area 5 (Kalaska et al. 1990), dorsal premotor cortex (Cisek et al. 2003), and some cells in primary motor cortex (Kakei et al. 1999; Scott and Kalaska 1997).

The key to understanding the intervening transformation may be the famous gaze-dependent “gain fields,” which multiply the overall response of neurons to visual input (Zipser and Andersen 1988). Gain fields responsive to both eye and head position are found through most of the structures discussed in the preceding paragraph (Batista et al. 1999; Battaglia-Mayer et al. 2000; Boussaoud et al. 1998; Brotchie et al. 1995; Mushiake et al. 1997), although less so as one approaches later stages closer to the motor cortex (Cisek et al. 2002). Gain fields appeared to lose their significance in the “direct transformation model” (Buneo et al. 2002), but in our scheme (Fig. 5), they become crucial in providing the possible substrate for rotating the gaze-centered hand motor error into shoulder coordinates. Gain fields could do this by tweaking the individual contributions of units so that the overall population vector rotates as illustrated in Fig. 5. Indeed, when neural nets were trained to perform the geometrically equivalent transformations for saccades (Smith and Crawford 2001b) this is exactly the solution that they arrived at. Thus smatterings of the complete solution to these transformations may already be visible in the known neurophysiological data.

Visual Control of Grasp

No discussion of eye–hand coordination would be complete without the inclusion of hand control itself. When we reach out to pick up an object, not only does our hand extend to the correct location, but our grasp adjusts its shape in anticipation of the target’s size and orientation well before contact is made. An efficient grasp requires the coding of an object’s spatial location and intrinsic properties (size and shape), and the transformation of these properties into a pattern of distal (finger and wrist) movements. Although the parietal cortex has long been considered a high-order sensory area, specialized for spatial awareness and the directing of action, its role in processing 3-D shape for grasping is now becoming clearer.

Adjacent to the medial intraparietal cortex (MIP, which corresponds closely to the functional area PRR discussed above) lies the anterior intraparietal cortex (AIP). AIP includes neurons that code the size, shape, and orientation of graspable objects such as rings, plates, and cylinders (Gallese et al. 1994; Murata et al. 1996, 2000; Taira et al. 1990). These features help determine the posture of the hand and fingers during a grasping movement. AIP cells are maximally activated when particular finger/hand postures are made, under visual guidance, toward target objects. AIP is thus concerned with the visual guidance of the hand movement, especially in matching the pattern of movement with the spatial characteristics of the object to be manipulated (Sakata et al. 1997).

This illustrates another way in which parietal cortex may deal not only with feedforward transformations (as described earlier) but also with moment-to-moment information about the location, structure, and orientation of objects in egocentric coordinates, and thereby mediate the visual control of reaching (Conolly et al. 2000; Desmurget et al. 1999; Snyder et al. 2000) and grasping (Gallese et al. 1994; Murata et al. 1996, 2000; Taira et al. 1990). Presumably, though, for the sake of speed and accuracy, such feedback-driven adjustments make use of comparisons with stored representations and the sort of accurate feedforward transformations described in preceding sections.

How does grasp interact with reach?

Previous research has suggested that all 3 conceptual components of a grasping movement (transportation, rotation, and opening of the hand) have access to a common visual representation of an object’s orientation (Mamassian 1997). Recently we investigated the contribution of upper and lower arm torsion to grasp orientation during a reaching and grasping movement (Marotta et al. 2003). As the required grasp orientation increased from horizontal to vertical, there was a significant clockwise torsional rotation in both the upper and lower arms. Thus it appears that the upper and lower arms, and fingers forming the grasp, all rotate in coordination with one another to achieve the torsion necessary to successfully orient the grasp. In contrast, the work space–dependent aspects of arm torsion in the reach were independent of grasp, resulting in a kind of kinematic constraint called Donders’ law: one arm orientation for each reach location and grasp orientation (Hore et al. 1992; Medendorp et al. 2000).

One possible explanation for this, consistent with the discussion in our previous sections here, is that parietal regions MIP and AIP encode higher level “goals,” like desired hand
location and grasp orientation, respectively, leaving the details of kinematics for downstream motor areas like primary motor cortex. This could mean that a higher-level grasp orientation command from parietal cortex (perhaps represented in AIP) is mapped onto a motor control system that implements the rules for coordinated grasp orienting, whereas the higher-level reach goals commands (perhaps represented in MIP) maps more directly onto the position-dependent rules that result in Donders’ law (Marotta et al. 2003).

CONCLUSIONS

Eye–hand coordination is complex, so any simplifying rule that might help us understand its neural underpinnings is potentially very useful. One such rule is that ultimately it is about visual guidance of the hand. Eye movements themselves do not directly affect the behavioral outcome, so they can be the slave to the master—the visually guided reach system. This simple fact helps to explain many of the disparate findings in the area of motor eye–hand strategies, and also points toward the importance of the neural mechanisms that compensate for the constant intervening motion of the eyes, motion that that might otherwise interfere with visually guided reaching. The process of spatial updating frees up the gaze control system to serve vision in more flexible ways.

This viewpoint also points us in the direction of understanding the neural mechanisms that map spatial vision onto the patterns of muscular contraction required for accurate reaching and grasping. These mechanisms are dauntingly complex in that they could potentially involve much of the brain, and many recurrent feedback loops. However, once again simplifying principles hold. The brain must implement certain fundamental transformations in a certain order, and it appears to do so in a modular fashion, where certain modules fit more closely with particular aspects of eye–hand coordination.

Despite the many recent advances in understanding the mechanisms of eye–hand coordination, numerous questions remain unanswered. For example, how are the feedforward transformations discussed in this review nested within both internal feedback loops (Battaglia-Mayer et al. 2003) and within loops involving visual feedback (Desmurget et al. 1999; Pisella et al. 2000)? How (at the detailed neural network level) are visual signals, proprioceptive signals from the hand, eye orientation signals, and head orientation signals compared in and beyond the PPC (Buneo et al. 2002; Crawford et al. 2000; Henriques and Crawford 2002)? Do gaze-centered PPC representations (Batista et al. 1999; Medendorp et al. 2003a,b) code the sensory targets for a movement, the desired movement itself, or something more abstract, like the goal of the movement? How do these signals interact for the purpose of selecting both targets and the effectors (e.g., right or left hand) that act on them (Carlton et al. 2002; Cisek et al. 2003)? And how do these transformations achieve the plasticity required for arbitrary stimulus-response associations, beyond simply reaching toward an object? Refining schemes like that shown in Fig. 5, and refining our knowledge of how these schemes relate to real neurophysiological signals, is one way to approach these questions.

GRANTS

Some work described in this review was funded by grants from the Canadian Institutes of Health Research (CIHR) and National Sciences and Engineering Council of Canada held by J. D. Crawford. J. D Crawford is supported by a Canada Research Chair. W. P. Medendorp is supported by the Human Frontier Science Program and the Netherlands Organization for Scientific Research. J. J. Marotta is supported by a Senior Research Fellowship granted by CIHR Institute of Neurosciences, Mental Health, and Addiction.


