Three-Dimensional Transformations for Goal-Directed Action

J. Douglas Crawford,1 Denise Y.P. Henriques,2 and W. Pieter Medendorp3

1York Centre for Vision Research, Canadian Action and Perception Network, and Departments of Psychology and 2Kinesiology and Health Sciences, York University, Toronto, Ontario, Canada, M3J 1P3; email: jdc@yorku.ca, deniseh@yorku.ca
3Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, 6525 HR, Nijmegen, The Netherlands; email: p.medendorp@donders.ru.nl

Abstract

Much of the central nervous system is involved in visuomotor transformations for goal-directed gaze and reach movements. These transformations are often described in terms of stimulus location, gaze fixation, and reach endpoints, as viewed through the lens of translational geometry. Here, we argue that the intrinsic (primarily rotational) 3-D geometry of the eye-head-reach systems determines the spatial relationship between extrinsic goals and effector commands, and therefore the required transformations. This approach provides a common theoretical framework for understanding both gaze and reach control. Combined with an assessment of the behavioral, neurophysiological, imaging, and neuropsychological literature, this framework leads us to conclude that (a) the internal representation and updating of visual goals are dominated by gaze-centered mechanisms, but (b) these representations must then be transformed as a function of eye and head orientation signals into effector-specific 3-D movement commands.

Keywords

vision, saccades, reach, reference frames, parietal cortex, brainstem
INTRODUCTION

Day-to-day life can be described as a series of goal-directed behaviors, sometimes relatively simple and direct, such as pressing a doorbell, and sometimes highly abstract through space and time, such as planning a university education. Here, we focus on the spatial transformations for two simple and well-studied behaviors: gaze and hand movements made immediately, or after a short delay, toward a visual goal. We consider how these systems represent spatial goals, how these systems update spatial goals during self-motion, and finally how they transform goals into action.

The major theme of this review is that internal representations and transformations, even for extrinsic goals, cannot be divorced from the underlying three-dimensional (3-D) geometry that links the sensors to the effectors. This geometry affects not only how stimuli project onto the sensory apparatus, but also how visual activation maps onto the correct pattern of effector commands. These mappings, or transformations, must account for the translational, and especially rotational, geometry of the eyes, head, and shoulder. It may seem tempting to ignore some of these details, but the brain has no such luxury. Here, we focus on how these details are incorporated into the feed-forward (open-loop) transformations for movement. Viewed from this perspective, the early spatial transformations for visually guided gaze and reach movements show several common principles.

Unless stated otherwise, the behavioral data referenced below pertain to observations that hold for both the human and the monkey. Animal models continue to advance the boundaries of known physiology in this field, but wherever possible, we emphasize recent advances in human systems neuroscience. But first, we provide the necessary background of mathematical and geometric concepts.

GEOMETRIC FOUNDATIONS

The Vocabulary of Spatial Transformations

Positions and movements are normally represented as vectors, which for our purposes can be loosely defined as 3-D arrows with a certain length and direction. The point at the tail of this arrow coincides with the zero vector, often called the reference position. To be meaningful, these must all be defined within some coordinate frame. The latter incorporates two concepts: A reference frame is some rigid body, useful for describing the relative location or
orientation of the body we want to represent. In neuroscience, reference frames are typically divided into two categories: egocentric reference frames, where a location is represented relative to some part of the body, such as the retina, eye, head, or torso, and allocentric reference frames, where a location is represented relative to an external object. In the case of motor control, one generally chooses the more stable insertion point of a set of muscles as the egocentric frame of reference. For example, the head is the logical frame for eye movement, and the torso is the logical frame for head and arm movement. One can then fix a set of coordinate axes within this frame and use some arbitrary unit along these axes to specify the components of the vector. These topics have been reviewed previously (Soechting & Flanders 1992), and rigorous definitions can be found in any linear algebra text. Sometimes there is confusion about the meaning of an “eye-centered” reference frame. A frame of reference could be both eye-centered in the sense that its directional coordinates are fixed with the rotating eye and head-centered in the sense that the ego center of these coordinates is located at some fixed point in the head. Therefore, we use the term gaze-centered to denote a directional coordinate system that rotates with the eye.

When discussing position/movement, it is important to distinguish between location/translation and orientation/rotation. These two types of position/motion have very different mathematical properties. Vectors representing the former commute (they add in any order), but the latter do not: The math of rotations is highly nonlinear and generally is influenced by the initial orientation. As we see in the next section, the early geometry of visuomotor control is dominated by orientation/rotation. And yet the vast majority of models that deal with this system use translational math that only approximates rotations over a small range. This principle was first noted in the context of oculomotor control (Tweed & Vilis 1987), but it has implications for nearly every process described below.

The 3-D Geometry of Visual-Motor Transformations

Gaze direction determines the two-dimensional (2-D) direction of the visual stimulus that falls on the fovea. However, the spatial correspondence of proximal stimuli on other points on the retina to the locations and the orientations of distal stimuli is determined by the complete 3-D orientation of the eye, including torsion. (Conversely, one can only infer the correct plan for a goal-directed movement from knowledge of both the proximal visual stimulus and 3-D eye orientation; see 3-D Reference Frame Transformations, below.) Here we define torsion as rotation about a head-fixed axis aligned with the primary gaze direction. Defined thus, Listing’s law (Figure 1a) states that torsion is held at zero (in practice, within ±1°). Mechanical factors likely play a role in implementing some aspects of Listing’s law (Demer 2006a, Ghasia & Angelaki 2005, Klier et al. 2006), but we know that they do not constrain torsion because Listing’s law is obeyed only for smooth pursuit and saccades with the head fixed (Ferman et al. 1987, Hashwanter et al. 1991, Tweed & Vilis 1987) and for gaze fixations during head translation (Angelaki et al. 2003). Other types of eye movement abandon or modify Listing’s law to optimize different factors such as retinal stabilization and binocular vision (Misslisch & Tweed 2001, Tweed 1997).

During natural gaze behaviors, subjects use a Fick strategy to move the head. This strategy implies that the head assumes orientations near zero torsion in Fick coordinates (Figure 1b), i.e., orientations that can be reached by a horizontal rotation about a body-fixed vertical axis and a vertical rotation about a head-fixed horizontal axis, with only minor, random variations about the third fronto-caudal torsional axis (Glenn & Vilis 1992, Klier et al. 2003, Medendorp et al. 1999).

In natural behavior, both the eyes and the head contribute to gaze direction, where the former contributes more vertical and the latter more horizontal (Freedman & Sparks 1997),
which results in a Fick-like constraint on eye-in-space orientation, i.e., with torsion minimized about the visual axis (Glenn & Vilis 1992, Klier et al. 2003, Radau et al. 1994). Because both the eye and the head each show random biological errors in torsional control during gaze, these errors sum to produce torsional variability of up to ±10°, a factor rarely accounted for in visual or motor experiments.
The eyes and shoulder joint are essentially capable of rotation only, but head motion and its visual consequences are more complex (Figure 1). Because the spine attaches near the back of the head and the eyes are near the front, any head rotation causes the two eyes to translate in different directions relative to space (Crane et al. 1997, Crane & Demer 1997, Medendorp et al. 2000) and the shoulder (Henriques et al. 2003, Henriques & Crawford 2002). Separation of the eyes is crucial for stereoscopic vision, but it also provides two different head-centered reference locations for visual direction. Psychophysical experiments suggest that visual direction is aligned to each eye independently (Erkelens & van de Grind 1994), to a central cyclopean eye (Ono et al. 2002), or to a dominant eye (Porac & Coren 1976), likely depending on the task. In visuomotor tasks that encourage monocular alignment, dominance may switch, depending on the field of view (Banks et al. 2004, Khan & Crawford 2001).

Goal Coding versus Sensory and Motor Coding
High-level goal representations are closely associated with working memory and the dissociation of future intentions from current sensorimotor events (di Pellegrino & Wise 1993, Goldman-Rakic 1992). Here, we restrict this notion to entail early visuomotor representations of desired gaze and hand positions. If these encode spatial goals, one should be able to discriminate this activity from both sensory and motor events. Anti-saccade (or anti-reach) tasks dissociate the direction of the visual stimulus from the direction of the internal goal for movement (Guitton et al. 1985, Munoz & Everling 2004). Subjects are trained or asked to move in the direction opposite of the stimulus (pro-saccades/pro-reaches refer to movements made directly to the target). Recordings from monkey LIP and MIP during anti tasks suggest that most neurons are tuned for the movement direction, some encode the visual stimulus direction, and some switch from the latter to the former during the trial (Gail & Andersen 2006, Hallett 1978, Kusunoki et al. 2000, Zhang & Barash 2000). Similarly, human PPC is spatially selective for direction in prosaccades/reaches and remaps this activity to tune to the opposite direction during anti tasks (Medendorp et al. 2005, Van Der Werf et al. 2008). However, when subjects were instead trained to point while looking through left-right reversing prisms, the spatially selective activity in most PPC areas [superior parieto-occipital cortex (SPOC), mIPS, visual areas V3, 7] remained tied to the visual direction of the goal, not the movement direction (Fernandez-Ruiz et al. 2007). Only one PPC region—the angular gyrus (Figure 2)—showed the opposite effect. Taken together, these experiments suggest that visuomotor areas such as SPOC primarily code the spatial goal for movements.

Egocentric versus Allocentric Coding
The dorsal stream of vision (terminating in parietal-frontal movement areas) is, by default,
Figure 2
Overview of visuomotor brain areas and transformations. (a) Schematic representation of human brain (lateral view) regions involved in processing of visuomotor transformations and eye-hand coordination: VC, visual cortex (V3A); AG, angular gyrus; mIPS, mid-posterior intraparietal sulcus; and SPOC, superior parieto-occipital cortex; S1, primary somatosensory area for arm movements (proprioception); BA5, Brodmann area 5; M1, primary motor cortex; PMd, dorsal premotor cortex; FEF, frontal eye fields; SEF, supplementary eye fields; DLPC, dorsolateral prefrontal cortex; SC, superior colliculus; PCS, precentral sulcus; CS, central sulcus; IPS, intraparietal sulcus; POS, parieto-occipital sulcus. (b) Primarily eye-centered ego(centric) goal representations interact with allo(centric) representations and are updated as a function of eye rotation. These signals are then put through an inverse internal model of the eye-head-torso system to compute motor effector commands for limb and gaze control. Efference copies derived from the latter provide position and movement signals for the internal model and updating, respectively, whereas hand position signals derived from multiple sources are used in computations of the reach command (see text for details).
involved in egocentric coding, i.e., relative to some part of the body (Goodale & Milner 1992, Schenk 2006). Neurophysiological studies have attempted to determine the frame of reference by dissociating the candidate frames (most often the eye and head) while recording sensorimotor receptive fields. With some exceptions (e.g., Avillac et al. 2005, Mullette-Gillman et al. 2009), most studies of goal-related activity in PPC, frontal cortex sac- cade areas, and superior colliculus (SC) suggest a gaze-centered, eye-fixed frame of reference (Andersen & Buneo 2002, Colby & Goldberg 1999). Consistent with this, fMRI recordings show egocentric directional tuning over human parietal and frontal visuomotor areas (Kastner et al. 2007, Levy et al. 2007, Medendorp et al. 2006, Schluppeck et al. 2005, Sereno et al. 2001), with gaze-centered coding in PPC and dorsal premotor cortex and body-centered coding for reaching near motor cortex (Beurze et al. 2010). However, this scheme may depend on the sensory modality used to aim the action: When the goal stimulus is somatosensory, PPC seems capable of switching from gaze-centered to body-centered coordinates (Bernier & Grafton 2010).

In contrast, the ventral visual stream (including occipital-temporal areas involved in object recognition) is more closely associated with allocentric coding, i.e., relative to some stable external visual cue (Goodale & Milner 1992, Schenk 2006). The brain likely relies more on these mechanisms when memory delays increase (Glover & Dixon 2004, Goodale & Haffenden 1998, Obhi & Goodale 2005), perhaps because allocentric codes are more stable over time (Carrozzo et al. 2002, Lemay et al. 2004, McIntyre et al. 1998). However, to influence behavior, allocentric signals must somehow enter the action stream (Figure 2b). Consistent with this notion, egocentric codes appear in visual area 7a before allocentric codes do (Crowe et al. 2008). Monkeys trained to saccade toward a particular end of an object show object-centered spatial tuning in supplementary eye fields (SEF), area 7a and LIP (Olson & Gettner 1996, Olson & Tremblay 2000, Sabes et al. 2002, Tremblay & Tremblay 2002). However, these areas may use objects as a reference position, whereas the underlying reference frame may still be egocentric. For example, Deneve & Pouget (2003) showed, with the use of neural network models, that object-centered spatial tuning can arise from neurons with gaze-centered receptive fields that show object-modulated firing rates.

When both egocentric and allocentric cues are available, the brain uses both (Battaglia-Mayer et al. 2003, Diedrichsen et al. 2004, Sheth & Shimojo 2004), incorporating allocentric information at least until movement begins (Hay & Redon 2006, Krigolson & Heath 2004). Allocentric and egocentric cues are combined on the basis of both actual reliability and subjective judgments of their relative reliability (Byrne & Crawford 2010).

Spatial Updating: Behavioral Aspects

Animals are not always motionless when planning goal-directed movements. Often self-motion invalidates the spatial relationship between extrinsic stimuli and the intrinsic sensory representations they produced. One option would be to wait for new sensory feedback, but this would introduce processing delays (e.g., the duration of a saccade + the latency for visual feedback) that could at times mean the difference between life and death. Moreover, this combined latency (~200 ms) multiplied by 3–4 saccades/second would render us functionally blind during most of our waking lives. To avoid such delays and blind periods, the brain must derive a predictive representation of visual space from brief visual glimpses and copies of motor commands (Ariff et al. 2002, Desmurget & Grafton 2000, Mehta & Schaal 2002, Wolpert & Ghahramani 2000). The process that updates spatial presentations during self-generated or passively induced motion is called spatial updating.

Spatial updating is often studied in the double-step task, in which subjects view a target, produce an intervening eye movement, and then move toward the first target. Saccades can

The studies cited above were performed in dark conditions, forcing subjects to rely on their own egocentric sense of target direction. Visuomotor systems may use different strategies when visual feedback is available (Flanagan et al. 2008). However, even when visual feedback is available, humans are more accurate at aiming movements when spatially updated memory of the goal is also available (Vaziri et al. 2006).

Theoretical Mechanisms for Spatial Updating

As we have already seen, the early spatial representations for visual goals, from the retina to the PPC and some areas of frontal cortex, utilize primarily an eye-fixed, gaze-centered code. This code could be used in two general ways to provide spatial updating. First, it could be compared with eye, head, and even body position. For example, many of these same areas contain subtle eye-position modulations called gain fields (Andersen & Buneo 2002, Boussaoud & Bremmer 1999, Sahani & Dayan 2003) that could, in theory, transform gaze-centered signals into successively more stable frames such as the head or body (Zipser & Andersen 1988). The problem with this scheme is that (although motor commands are eventually encoded in effector-specific, muscle-based coordinates) there is little evidence for visuospatial representation in such intermediate spatial maps.

The alternative is to use the internal sense of self-motion to remap the goal representation within gaze-centered coordinates, so that after the eye movement it corresponds to the correct retinal location at final eye position (Colby & Goldberg 1999). This model was originally simulated by subtracting a vector representing the intervening eye movement from another vector representing the goal to obtain a third vector representing the final saccade direction in retinal coordinates (Moschovakis & Highstein 1994, Waitzman et al. 1991). This does not quite work in real-world conditions because (a) spatial updating of saccades is noncommutative (Klier et al. 2007, Smith & Crawford 2001), and (b) during torsional eye rotations, goals on the opposite side of gaze need to be updated in opposite directions (Crawford & Guitton 1997, Medendorp et al. 2002, Smith & Crawford 2001). However, the remapping model does work when the correct 3-D math is used. Neural network simulations show that these noncommutative operations can be performed through a combination of physiologically realistic eye orientation and movement commands (Keith & Crawford 2008).

Experimental Evidence for Remapping

Remapping occurs in virtually every area of the monkey brain associated with saccade and reach goal coding, including early visual areas (Nakamura & Colby 2002), LIP (Duhamel et al. 1992a, Gnadt & Andersen 1988, Heiser & Colby 2006), SEF (Russo & Bruce 2000), frontal eye fields (FEF) (Sommer & Wurtz 2008, Umeno & Goldberg 1997), MIP (Batista et al. 1999, Buneo et al. 2002), and the SC (Walker et al. 1995). Many neurons in these areas show peri-saccadic changes consistent with a recalculation of future saccade goals with respect to the new eye position, sometimes beginning even before the saccade (Duhamel et al. 1992a, Umeno & Goldberg 1997, Walker et al. 1995). Recent experiments suggest that
this is accomplished in part through signals routed from the brainstem, via the thalamus, to the cortex (Sommer & Wurtz 2008).

Gaze-centered remapping was first demonstrated in the human using a psychophysical paradigm (Henriques et al. 1998). This experiment relies on the control finding that humans overestimate the angle between a remembered peripheral pointing target and gaze direction (Bock 1986, McGuire & Sabes 2009). When subjects were additionally required to make a saccade between seeing, remembering, and pointing toward a central target (Figure 3a), the resulting pointing errors matched the final (updated) target-gaze angle, not the angle at the time of viewing. The same result occurred for pointing to targets at different distances (Medendorp & Crawford 2002), after body translations (Van Pelt & Medendorp 2007), after smooth-pursuit eye movements (Thompson & Henriques 2008), for pointing to goals inferred from expanding motion patterns (Poljac & van den Berg 2003), or proprioceptive and auditory targets (Jones & Henriques 2010, Pouget et al. 2002), and for repeated pointing movements to the same remembered target (Sorrento & Henriques 2008).

Human cortical remapping has been confirmed using several different approaches. fMRI recordings demonstrated that both remembered movement goals (Medendorp et al. 2003a) and passively remembered stimuli (Merriam et al. 2003, 2007) remap between the intraparietal sulci on opposite hemispheres during saccades (Figure 3b). Application of transcranial magnetic stimulation (TMS) pulses to the same cortical area disrupts remapping (Chang & Ro 2007, Morris et al. 2007). Unilateral optic ataxia patients show gaze-centered reach deficits that remap across saccades—from the “good” to “bad” hemifield, and vice versa (Khan et al. 2005b). Bidirectional saccadic updating was present in a patient with just one hemisphere (Herter & Guittton 1998) and recovered after anterior commissurotomy (Berman et al. 2005). Damage to frontal-parietal cortex can also produce deficits consistent with an impairment to the signal that drives updating (Duhamel et al. 1992b, Heide et al. 1995).

These findings do not show that gaze-centered remapping is the only mechanism for spatial updating. For example, patients with bilateral parietal-occipital damage appear to retain a different, nonretinal mechanism (Khan et al. 2005a). However, gaze-centered updating is likely the dominant mechanism for updating visual saccade and reach goals.

Encoding and Updating in Depth
In the previous section, we consider the encoding and updating of visual direction for action, but this leaves out an essential component: depth. Distance is a significant variable in the programming of vergence eye movements and reaching movement. It is generally assumed that target depth and direction are processed in functionally distinct visuomotor channels (Cumming & DeAngelis 2001, Vindras et al. 2005).

Depth perception is typically associated with binocular disparity. If one can correctly match, point-by-point, the images on the two retinas, then geometry dictates that they will be slightly deviated on the basis of the difference in distance of the target relative to the individual eyes, the interocular distance, and the 3-D orientations of the eye and head (Wei et al. 2003). The binocular version of Listing’s law partially reduces the degrees of freedom of this comparison (Tweed 1997), but in the absence of other visual cues, knowledge of 3-D eye and head orientation is required (Blohm et al. 2008). But these egocentric mechanisms are normally supplemented by allocentric cues based on object features and pictorial information, such as relative size, perspective, occlusion, and convergence of lines (Howard & Rogers 1995, Wei et al. 2003).

Spatial updating also occurs in depth, i.e., humans partially compensate for changes in vergence angle that occur between sensation and action (Krommenhoek & Van Gisbergen 1994). A recent study by Van Pelt & Medendorp (2008) used a variation of the
Psychophysical and neuroimaging evidence for gaze-centered spatial updating in the human. (a) Reaches to memorized visual targets, presented on the fovea, are relatively accurate (control trial), whereas reaches to peripheral targets show a clear directional bias (fixation trial). Reaches to a foveally presented target, but shifted to the periphery by an intervening saccade (saccade trial), show the same bias as do fixation trials, suggesting that the target is updated relative to gaze (Henriques et al. 1998). (b) A bilateral region in the PPC (red) shows gaze-centered spatial updating during the intervening saccade task. When eye movements reversed the side of the remembered target location relative to fixation, the region exchanged activity across the two cortical lobules (left-right trial). Modified from Medendorp et al. (2003a)
paradigm by Henriques et al. (1998) to show that similar principles hold for depth updating. Considering the binocular fixation point as the 3-D depth equivalent to the gaze point, these authors measured errors for reaching to targets relative to that depth. These results suggest that targets of reaching movements are updated in both direction and depth relative to the binocular fixation point (Van Pelt & Medendorp 2008).

In more complex motion conditions, direction and depth cannot be regarded as independent variables in the neural computations of spatial updating. In motion parallax, the change of visual direction depends on target depth during head translation. Psychophysical evidence shows that the brain takes this translation-depth geometry into account when programming the direction of saccades after an intervening translation, even compensating for eye translations produced by head rotation (Klier et al. 2008, Li et al. 2005, Li & Angelaki 2005, Medendorp et al. 2003b, Van Pelt & Medendorp 2007). To control such behavior in gaze-centered coordinates, the updater circuit must synthesize information about self motion with object depth information to remap each target by a different amount (Medendorp et al. 2003b).

TRANSFORMATION OF THE GOAL INTO A MOVEMENT COMMAND

Once a goal has been selected (Schall & Thompson 1999), and a desired action chosen (Cisek & Kalaska 2010), the representations described in the previous sections must be transformed into commands suitable for action (Figure 2b). In real-world circumstances, this transformation would be combined with visual feedback (Gomi 2008), but here we focus on the feed-forward mechanisms required for rapid, accurate action.

Computing the Displacement Vector

In theory, motor systems could function by specifying desired postural patterns and letting the effector drift to that position (Bizzi et al. 1984, Feldman 1986). However, physiological experiments suggest that early saccade and reach areas are concerned primarily with developing a plan to displace gaze and/or hand position.

Retinal stimulation defines a desired gaze displacement, implicitly relative to current gaze direction, in eye-fixed coordinates. Subsequent oculomotor codes maintain this gaze-centered organization, computing eye velocity and orientation commands only at the final premotor stage before motoneurons (Robinson 1975). The exception occurs for depth saccades, in which current and desired binocular fixation must be compared to program a disconjugate saccade component. Saccade-related neurons in LIP show modulations related to both initial and desired depth (e.g., Genovesio et al. 2007). A fixed relationship rarely exists between initial hand position, the goal, and gaze direction. The only way to compute the reach vector is to compare initial and desired hand position. For translational motion of the hand, it is sufficient to subtract a vector representing initial hand position from a vector representing desired hand position in the same frame. Investigators have historically assumed that this was done either entirely in visual coordinates or by transforming the visual goal into proprioceptive coordinates.

Sober & Sabes (2005) showed that when vision is available, humans compare the target to both visual and proprioceptive sensation of hand position and optimally integrate these signals depending on the stage of motor planning; however, they tend to rely more on vision especially in the early stages of motor planning. Other psychophysical experiments in healthy and brain-damaged humans have supported the notion that the reach vector is calculated either in gaze-centered coordinates (Chang et al. 2009, Khan et al. 2005b, Pisella et al. 2009, Pisella & Mattingley 2004) or in a mix of gaze and somatosensory coordinates (Beurze et al. 2006; Blangero et al. 2007; Khan et al. 2005a,b).

Several regions within PPC (Figure 2) play a role in both visual and proprioceptive calculations of the 3-D reach vector. Parietal area 5
is modulated both by target depth signals and by initial hand position (Ferraina et al. 2009b). Human angular gyrus appears to play a special role for incorporating the somatosensory sense of hand position into the reach vector (Vesia et al. 2010). Moreover, the PPC appears to possess the signals necessary for computation of movements in depth (see Ferraina et al. 2009a for review). Many neurons in areas such as LIP and the parietal reach region (PRR) are sensitive to both visual direction and retinal disparity (Bhattacharyya et al. 2009, Genovesio & Ferraina 2004, Gnadt & Mays 1995). Activity in most of these neurons is also modulated by vergence angle (Bhattacharyya et al. 2009, Genovesio & Ferraina 2004, Sakata et al. 1980). Consistent with these findings, damage to PPC produces deficits in both reach direction and depth (Baylis & Baylis 2001; Khan et al. 2005a,b; Striemer et al. 2009).

Buneo et al. (2002) showed that neurons in monkey PPC (area 5 and PRR) can show gaze-centered responses with hand-position modulations, consistent with calculation of the movement vector in visual coordinates. These responses persisted even when hand was not visible, suggesting that proprioceptively derived estimates had been transformed into gaze-centered coordinates. Recently, Beurze et al. (2010) reported similar findings in the human brain using fMRI. Other experiments suggest that dorsal premotor cortex (PMd) and PRR neurons show a relative-position code for target, gaze, and hand position (Pesaran et al. 2006, 2010) and/or encode target position in gaze coordinates with opposing gain modulations for gaze and hand position (Chang et al. 2009).

Reference frame transformations have historically been considered from the viewpoint of position coding, where retinal position is compared with eye position to compute target position relative to the head, and this is compared to head position to compute target position relative to the body (Flanders et al. 1992). For relative position/displacement codes, the need for such comparisons disappears in frames that only translate with respect to each other (Andersen & Buneo 2002, Goldberg & Colby 1992). However, the frames of reference for visuomotor transformations (eye, head, torso) primarily rotate with respect to each other (Figure 1). The mathematics of rotations dictates that the representation of a movement or position in one of these frames corresponds to different representations in the other frames as a complex, nonlinear product of their relative orientations (Blohm & Crawford 2007, Crawford & Guittion 1997). Small movement and position vectors restricted to a frontal plane (like a laboratory stimulus screen) are relatively immune to the effects, but this result does not hold in general real-world conditions. For example, if gaze is simply directed 90° to the left, a forward reach in body coordinates is now a rightward reach in eye coordinates. In most circumstances, these reference frame projections produce more complex distortions in gaze (Figure 4a) and reach (Figure 4b) space. In a system that relies on relative position/displacement codes, these nonlinearities become the central problem in reference frame transformations, and this can be solved only by a transformation that includes a model of eye/head orientations and rotational geometry.

A brain that does not account for this geometry would produce predictable errors in generating rapid movements (Crawford & Guittion 1997). Behavioral studies in humans have shown that saccades to visual targets partially account for torsional eye orientations and fully account for eye positions in Listing’s plane (Klier & Crawford 1998). A recent study has also shown that smooth-pursuit eye movements compensate for these factors (Blohm et al. 2006, Daye et al. 2010).
Pointing movements toward horizontally displaced targets also compensate for geometric relationships related to vertical eye position and the way it distorts the retinal projection (Crawford et al. 2000). Moreover, the internal models for reach and pointing movements also account for the translational linkage geometry (Figure 1c) between the centers of rotation of the eye, head, and shoulder (Henriques et al. 2003, Henriques & Crawford 2002).

A recent study (Blohm & Crawford 2007) combined all these features, modeling visually guided reach with the use of a direct transformation from visual coordinates to shoulder coordinates, accounting for only the translational geometry of the system, versus a system with a full internal model of eye-head-shoulder linkage (Figure 1) and nonlinear reference frame transformations (Figure 4b). As expected, the former model predicted errors in both reach direction and depth as a function of initial eye orientation, whereas the latter model predicted perfect reach. Tested the same way, real reaches showed various unrelated offsets and noise in the absence of visual feedback, but they did not show any of the errors predicted by the direct transformation model, even in the initial stages before proprioceptive feedback could occur.

3-D Reference Frame Transformations—Neural Mechanisms

The best theoretical candidate for reference frame transformations in the brain arises from studies of gain fields and their variants (Blohm et al. 2009, Pouget et al. 2002, Zipser & Andersen 1988). As mentioned above, these describe postural modulations (such as eye position) on visual-motor receptive fields. Eye and head position gain fields have been identified in essentially every area of the brain implicated in visuomotor transformations, from occipital cortex (Galletti & Battaglini 1989, Weyand & Malpeli 1993), to parietal eye and reach fields (Andersen & Mountcastle 1983, Brotchie et al. 1995, Chang & Snyder 2010, Galletti...
et al. 1995), to frontal cortex gaze and reach control centers (Boussaoud & Bremmer 1999, Boussaoud et al. 1998), and even subcortical structures (Groh & Sparks 1996a,b; Van Opstal et al. 1995). The original account of gain fields assumed the use of visual goal-in-space code (Zipser & Andersen 1988), which has since been questioned (Colby & Goldberg 1999), but the nonlinear geometry described in the last section gives new significance to this theory.

Artificial neural networks can be trained to transform visual targets into saccades (Smith & Crawford 2005) or reach movements (Blohm et al. 2009) using the correct 3-D geometry (Figures 1 and 4). These networks develop intermediate units that show gain fields similar to those seen in real physiology. Moreover, when probed with simulated receptive field mapping and microstimulation, individual units can show both a sensor-fixed frame of reference for the former and effector-fixed frames for the latter. This shows that (a) unit recording and stimulation reveal different neuron properties, and (b) individual units should show a fixed input-output relation when they perform a transformation.

These modeling studies suggest that electrical microstimulation reveals the reference frame to which a neural structure projects and should differ from the input code (derived from receptive field mapping) when a transformation is occurring. A 3-D reference frame analysis on gaze shifts evoked from the SC (Klier et al. 2001) and LIP (Constantin et al. 2007) showed that their position dependencies simply arise from the projection of light on an eye-fixed spherical frame. But these results also suggest that the 3-D reference frame transformation for gaze saccades occurs only as late as at the level of the brain stem/cerebellum. SEF stimulation evoked gaze shifts toward intermediate, eye-, head-, and body-fixed frames (Martinez-Trujillo et al. 2003b, Park et al. 2006), suggesting a capacity for more complex and arbitrary reference frame transformations in the frontal cortex.

The analogous 3-D analysis has not been done for reach, but as we have seen, superior parietal structures appear to encode primarily visual targets with a gaze-centered code (Batista et al. 1999, Bhattacharyya et al. 2009), intermediate structures such as angular gyrus (Fernandez-Ruiz et al. 2007, Vesia et al. 2010) and ventral premotor cortex (Beurze et al. 2010, Kakei et al. 2001) employ progressively more extrinsic reach codes, and structures closer to the motor output for reach employ successively more effector-related spatial codes (Beurze et al. 2010, Hoshi & Tanji 2004, Scott 2003). Moreover, the latter structures continue to encode gaze-fixed signals (Boussaoud & Bremmer 1999, Cisek & Kalaska 2002) and yet produce complex and coordinated movements when stimulated (Graziano et al. 2002a,b). This seeming paradox could reflect a transition from sensory to motor codes such as that seen in 3-D network models (Blohm et al. 2009).

The 2-D to 3-D Transformation

Finally, the lower-dimensional neural codes discussed in the previous sections must be converted into the commands that implement the higher-dimensional behavioral geometry shown in Figure 1. The mechanisms that convert 2-D gaze commands into 3-D eye rotations and implement Listing’s law have been the subject of intense theoretical debate (e.g., Quaia et al. 1998, Quaia & Optican 1998, Raphan 1998, Tweed & Vilis 1987). The analogous transformations for reach have also been modeled (Lieberman et al. 2006).

High-level gaze-control centers (SC, FEF, SEF) appear to encode the desired 2-D direction of gaze, leaving 3-D eye and head control downstream (Monteon et al. 2010; Hepp et al. 1993; Klier et al. 2003; Martinez-Trujillo et al. 2003a,b; van Opstal et al. 1991). In contrast, the reticular formation saccade generator (Henn et al. 1989, Luschei & Fuchs 1972) and the neural integrator that holds eye and head orientation (Crawford et al. 1991, Fukushima 1991, Helmchen et al. 1996, Klier et al. 2002) utilize a 3-D coordinate system. Thus, a 2-D to 3-D transformation must occur between these stages.
The default 2-D to 3-D transformation cooperates with mechanical factors to maintain eye and head orientation within the Listing and Fick ranges (Figure 1). The brainstem coordinates for 3-D eye control are organized such that they effectively collapse into 2-D axes in Listing’s plane during symmetric bilateral activation of the midbrain (Crawford 1994, Crawford & Vilis 1992). The position-dependent torsional saccade axis tilts required (counterintuitively) to keep eye position in Listing’s plane (Tweed et al. 1990, Tweed & Vilis 1987) are then implemented mechanically (Ghasia & Angelaki 2005; Klier et al. 2006, 2011), possibly by pulley-like actions of tissues surrounding the eye muscles (Demer 2006a,b).

Similar neuro-muscular principles hold for head control: the 3-D brainstem coordinates for head control align with Fick coordinates (Klier et al. 2007), but neck anatomy also facilitates head rotations in Fick coordinates (Graf et al. 1995).

Different neural mechanisms are required to generate torsional movements toward or away from these 2-D ranges, for example during head-unimmobilized gaze shifts (Klier et al. 2003). A bilateral imbalance of input to the midbrain coordinate system is required to produce torsional components. What chooses the correct level of torsion? The cerebellum may influence torsional control in both the vestibular system, via outputs to vestibular eye-head cells with 3-D properties (Ghasia et al. 2008), and the saccade system, via inputs from the paramedian pontine reticular formation (Van Opstal et al. 1995). Consistent with this, it has been observed that Listing’s plane is degraded in patients with damage to the cerebellum (Straumann et al. 2000).

Analogous neural mechanisms may come into play for 3-D reach constraints, but these are less understood at this time.

**CONCLUSIONS**

Neural recordings from the human and monkey suggest that gaze and reach movements toward visual goals are controlled by separate, but overlapping neural control systems. When considered from the perspective of the 3-D geometry of the spatial relationship between goal representations and effector commands, and the associated computational problems that must be solved, these two systems show several common principles (Figure 2b). First, their early representational phases are dominated by gaze-centered mechanisms (although these coexist with other mechanisms, both egocentric and allocentric). Second, these gaze-centered signals are remapped during self motion. Third, upon selection for potential action, these representations are put through a series of transformations, involving computation of the movement vector (for depth saccades and reach), a successive series of reference frame transformations, and finally elaboration of these higher-level/low-dimensional plans into multidimensional motor commands. The role of some of these stages and their corresponding physiology—such as the prevalence of eye-position signals throughout the visuomotor system—becomes fully clear only when one takes the complete 3-D geometry of the system into account. Given the commonalities that emerge in these two systems, one would expect similar physiological solutions to arise whenever other sensorimotor systems encounter similar computational problems.

**DISCLOSURE STATEMENT**

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

**ACKNOWLEDGMENTS**

The authors thank Dr. Luc Selen for help with preparing Figures 1b and 1c, 3, and 4b, and Dr. Michael Vesia for help preparing Figure 2a. Dr. Crawford’s work was supported by a Canadian Institutes of Health Research (CIHR) Canada Research Chair and grants from CIHR.
the National Science and Engineering Research Council (NSERC) and the Canada Foundation for Innovation (CFI). Dr. Henriques’s work was supported by grants from NSERC, CFI, the Ministry of Research and Innovation (Early Researcher Award), and the J.P. Bickell Foundation. Dr. Henriques is an Alfred P. Sloan fellow. Dr. Medendorp’s work was supported by grants from the Human Frontier Science Program (HFSP), the Netherlands Organisation for Scientific Research (NWO), and the Donders Center for Cognition.

LITERATURE CITED


Jones SA, Henriques DY. 2010. Memory for proprioceptive and multisensory targets is partially coded relative to gaze. *Neuropsychologia* 48:782–92


Mays LE, Sparks DL. 1980. Saccades are spatially, not retinocentrically, coded. Science 208:1163–65
Schlag J, Schlag-Rey M, Dassonville P. 1990. Saccades can be aimed at the spatial location of targets flashed during pursuit. *J. Neurophysiol.* 64:575–81


