Eye-centered vs body-centered reaching control: A robotics insight into the neuroscience debate

Tran Minh Tuan, Philippe Souères, Michel Taïx
CNRS; LAAS ; 7, av. du Colonel Roche, 31077 Toulouse, France
Université de Toulouse; UPS, INSA, INP, ISAE ; LAAS ;
F-31077 Toulouse, France
E-mail: {tmtuan, soueres, taix}@laas.fr

Benoît Girard
ISIR, UPMC/CNRS, 4 place Jussieu,
University Paris 6,
F-75252 Paris, France
E-mail: benoit.girard@isir.fr

Abstract—Whether the central nervous system of primates uses a body-centered or an eye-centered frame for reaching is still a controversial debate in neurosciences. Does the proprioceptive information allow to represent the visual information with respect to the body or inversely, does it transform the hand position in retinal coordinates? In this paper, we implement and test two control schemes associated to each of these two hypotheses by using a computational approach from robotics. To this end, biological models of motor control are applied to a realistic dynamic model of upper body including a 4 DOF eye-neck kinematic chain and a 6DOF arm. Important characteristics, such as proprioceptive biases and sensory delay, are considered. Simulation results allow to compare the geometry of trajectories and illustrate the better robustness of the eye-centered control scheme with respect to biases and sensory delay. Applications to the control of the humanoid robot HRP2 are finally presented.

I. INTRODUCTION

To plan and control visually guided reaching movements, the Central Nervous System (CNS) of primates needs to perform various transformations in order to interface sensory data and motor variables which are encoded in different referentials. The image of the object, initially formed in the retina, must be related to the hand location via the combination of visual and proprioceptive information to allow the synthesis of a consistent motor input. Different experiments tend to prove that the reference input used by the CNS for planning the motion is the position error – called “difference vector” – between the hand and the target [1]. In order to make possible the coding of this different vector, the hand and the target need to be expressed with respect to the same reference frame. The question of which spatial reference is used by the CNS to encode this vector is a controversial debate in neurosciences.

A first way of thinking is that the CNS encodes the difference vector with respect to the body. This idea was suggested by early experiments showing that, during motion, the activity of different cortical areas varies as a function of the end-effector location with respect to the body. At the spinal level and in S1, the activity of cells seem to reveal a body centered representation of hand direction [2]. It was shown that during the delay period of reaching experiments in monkeys, cells in M1, PMd and 5d [3] are tuned to the direction of the upcoming movement. Georgopoulos et al. proposed a model in which population of neurons in M1 encode preferred direction of motion with respect to the body [4]. If the CNS uses a body-centered frame to encode the hand-target difference vector then, the target location, initially encoded in retinal coordinates, must be expressed with respect to the body (hand, shoulder,...). This transformation, which is based on the integration of proprioceptive (eye position in the orbit and head position relative to the trunk) and visual (target location on the retina) data, is likely to be performed in the Posterior Parietal Cortex (PPC).

Today, a large number of neuroscientists share another way of thinking, which is that the CNS rather encodes the hand-target difference vector in visual coordinates. According to this idea, the activity of PPC, is interpreted in a different way. Instead of transforming the target location from eye-centered to body-centered coordinates, neural networks in this area should integrate proprioception and vision to express the hand location in eye-centered coordinates. Cohen and Andersen [5] suggested the existence of a common eye-centered reference frame for movements plans. It was shown that this eye-centered spatial representation constitutes a reference for multisensory integration, illustrating the predominant role of vision in primates [6]. Buneo et al. [7] observed that, when the relative hand-target position remains constant with respect to the fixation point, the activity of area 5d cells varied little, suggesting that the difference vector is encoded in fixation-centered coordinates. Finally, numerous psychophysics studies have strengthened the idea that visually guided movements are planned in an eye-centered frame [8]. However, though most models of reaching motor control deal with arm dynamics, very few consider the influence of eye and head rotations explicitly. Indeed, most studies which aim at determining which cost criterion is minimized by the CNS during movement suppose implicitly that the location of the target with respect to the body is known [9]. It has been argued by supporters of the eye-centered side that, in first approximation, changes in gaze direction due to eye and/or head rotations do not modify the hand-target difference vector in fixation centered coordinates. However, experiments showing that the activity of PMd and PMv neurons during reaching is modulated by gaze direction weaken this assertion [10].

In this paper we propose to evaluate the effect of sensory transformations related to the eye, neck and arm joints in both body-centered and the eye-centered control schemes by using a computational approach from robotics. Clearly, our objective is not to propose a model more biologically plausible than existing ones, but rather to compare the architectures of body-
centered and the eye-centered control schemes by using the robotics formalism. For both schemes, the same well-tested biological principles of motor control will be applied in order to reproduce the main characteristics of human movement while dealing with redundancy, proprioceptive biases, sensory delay, by considering the whole dynamics of the arm and the eye-neck links. Our aim is to provide computational arguments to answer the following questions: 1/ Does the transformation of the hand-target difference vector in a body-centered frame constitute a necessary preliminary step for arm control? 2/ Does expressing the difference vector with respect to the body, or the eye, induce variations of the hand trajectory? 3/ Is it possible to find computational arguments supporting the fact that one of these two control schemes is more robust with respect to biases and sensory delays than the other?

II. CONTROL SCHEMES

In this section we present two plausible models of control architectures that could be used by the brain for reaching a visual target. Each control scheme corresponds to a different coding of the hand-target difference vector. In the first scheme, that will be called “body-centered” (Fig. 1) the difference vector is expressed with respect to the body, whereas in the second scheme, called “eye-centered” (Fig. 2), the difference vector is expressed in the eye frame. Both schemes include two control loops that allow to control the eye-neck kinematic chain (eye loop) and the arm kinematic chains (hand loop) in parallel. At the beginning of the task, the target may be located in the periphery of the visual field. Considering the retinal azimuth and elevation errors as input, the eye control loop will drive the eye-neck joints in order to bring the target into the fovea. The arm loop takes the hand position and the target position (expressed in the same reference frame) as input, and outputs the angular displacement of the arm that allows to bring the hand to the target location. In both schemes, the “direct model” blocks represent an internal computation of direct kinematics, based on proprioception. In the body-centered diagram, the direct model 1 computes the position of the target and the hand with respect to the body from the corresponding positions in eye-centered frame. In this computation, the estimated value of the current head-eye configuration is taken into account. The direct model 2 computes the hand position in body-centered frame directly from the proprioception of the arm. Finally, both evaluations of the hand position (by vision and proprioception) are aligned to provide an estimated value. In the eye-centered scheme, the direct model evaluates the hand position with respect to the eye by using proprioceptive measurement of the eye-neck and arm joints. This proprioceptive information is then aligned with the visual one to provide an estimated value. The next sections describe in details the eye and arm controllers that are used in both schemes.

A. Eye controller

In the mammal brain, eye movements are driven by four motor brainstem oculomotor nuclei (BON), dedicated to upward, downward, rightward and leftward movements. Antagonist BON inhibit each other, while horizontal and vertical movements are controlled in parallel, thanks to the eye pulleys properties [11]. A target can be fixated using saccades (fast eye-movements without visual feedback) and smooth pursuit eye movements (slow movements based on visual feedback). Each of these movements is processed by cortical and subcortical pathways sharing a large part of their functional architecture [12], which all project to the BON. Many crucial premotor and motor regions within these pathways are structured in topological maps encoding the position of targets in the visual field, using retino-centered reference frames. Based on these considerations, a model of spherical sensor can be considered for designing a neuro-inspired eye controller with the visual servoing formalism [13]. The target position is represented by the azimuth $\alpha$ and elevation $\beta$, with respect to the retinal frame (Fig. 3 right). In order to express the interaction matrix for the visual error $e = (\alpha, \beta)^T$, we combined the well-known optical flow equations [13], which expresses the link between the velocity of a point $(X, Y)$ on the planar image of a stenope...
model of camera, and its kinematic screw $^1T$ (see Fig 3 left), with the following coordinate transformation:

$$\begin{cases}
\dot{\alpha} &= -\frac{1}{1 + \tan2\rho} X \\
\dot{\beta} &= \cos\alpha \cos\beta[\sin\alpha \sin\beta X - \cos\beta Y]
\end{cases}$$

From these equations, the optical flow equation for a spherical camera can be expressed in terms of the retinal azimuth and elevation angles as follows:

$$\dot{e} = (\dot{\alpha}, \dot{\beta})' = L_{\alpha,\beta} T$$

where the interaction matrix $L_{\alpha,\beta}$ is defined by

$$L_{\alpha,\beta} =\begin{pmatrix}
\frac{-\cos\alpha \sin\beta}{\rho} & 0 & \frac{\sin\alpha \sin\beta}{\rho} & \frac{-\sin\beta}{\rho} & -1 & -\tan\beta \cos\alpha \\
\cos\beta & \frac{\cos\alpha}{\rho} & \frac{\sin\alpha}{\rho} & 0 & \sin\alpha & 0
\end{pmatrix}$$

A visual servoing scheme can be used for controlling the gaze direction. Considering that the eye has two degrees of freedom in rotation (pan-tilt), and that the neck provides two additional joints$^2$, the eye-neck kinematic chain that allows to orient the gaze can be represented by a 4-dimensional vector of joints $q_E$. Using the Jacobian $J(q_E)$ of this chain, the eye kinematic screw $T$ can be related to the joint velocity $\dot{q}_E$:

$$T = J(q_E) \dot{q}_E$$

By composing (1) and (3), and specifying a decay rate $\Delta\varepsilon$ for $e = (\alpha, \beta)'$, a kinematic controller can be expressed by:

$$\Delta q_E = (L_{\alpha,\beta} J(q_E))^+ \Delta\varepsilon,$$

where $^+$ stands for the pseudo inverse operator. This controller was implemented both in the body-centered and the eye-centered control architectures to control independently the gaze direction in order to foveate the target ($e = 0$).

**B. Arm controller**

The same arm controller was implemented in both schemes. At each time, the hand and the target locations are estimated and the difference vector $x_{dv}$ is computed. This vector constitutes the input of the arm controller and the output is the joint displacement (see Fig. 4). We used a variation of the Hoff and Arbib model [14] to compute the next position, velocity and acceleration of the hand at each time, in order to satisfy the main characteristics of the human arm movement. On this basis, the joint displacement that allows to perform the prescribed hand displacement is computed by minimizing a cost criterion related to the arm dynamics. This step allows to cope with actuation redundancy at each time. The next paragraphs describe this process with details.

1) **Instantaneous hand displacement:** Basically Hoff-Arbib’s model [14] can be considered as a feedback version of the minimum-jerk model [15], which allows to take into account perturbations (variation of target position). Considering that the estimate of the target location $x_f$ is the input, and defining the state of the system at time $t$ by $x_t = [x_t, v_t, a_t]'$, where $x_t$ is the hand location, $v_t$ its velocity and $a_t$ its acceleration, this model allows to compute the expected state at the next time step $x_{t+\Delta t}$. According to their model, the state variables must obey the following dynamics:

$$x_{t+\Delta t} = A x_t + B x_f$$

where:

$$A = \begin{bmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
-60/D^3 & -36/D^2 & -9/D
\end{bmatrix},\quad B = \begin{bmatrix}
0 \\
0 \\
60/D^3
\end{bmatrix},$$

$D = t_f - t$ is the remaining time duration to final time $t_f$.

In this paper we propose a variation of this model by considering $x_t$ as the norm of the difference vector $x_{dv}$ at time $t$, and $v_t$ and $a_t$ as its first and second order time derivative$^3$. The objective is to make $||x_{dv}||$ converges to 0 in order to drive the hand to the target. The state values computed at the next time step are used to determine the hand displacement $\Delta X$ in Cartesian coordinates from the relation:

$$\Delta X = v_{t+\Delta t}(x_{dv}/||x_{dv}||)\Delta t$$

This novel usage of Hoff-Arbib’s model makes the computation simpler and faster because it is realized only on a one dimension variable (the norm of the difference vector) instead of the 3D hand coordinates. It is also consistent with the fact that aimed hand movements are planned vectorially, i.e., in terms of distance and direction, rather than absolute position in space [16], [1]. As the difference vector is computed from the hand and target positions, we used the same procedure as Hoff and Arbib for integrating the latency of vision $\Delta L$ and proprioception $\Delta t$ in the feedback loop. The reasoning is based on the use of a sliding window integrator whose output is the integral of its input over a temporal delay (see [14]).

$^1$The kinematics screw is defined by $T = (v, w)'$, where $v$ and $w$ represent the linear and angular velocity of the camera respectively.

$^2$The roll angle, which is not preponderant in this study, was set to zero in order to simplify the computation of the Jacobian.

$^3$As the sign of the difference vector never changes, the time derivative of its norm is always well defined.
2) Instantaneous joint displacement: From $\Delta X$ given by (5), the next step is to compute the joint displacement $\Delta q$ that makes the hand execute the prescribed movement. As the arm configuration has 6 DOF, whereas the reaching task only constrains 3 DOF (target position), the system is redundant. Numerous authors agree that the CNS solve this problem by minimizing a cost criterion. Such an optimization problem can be stated as follows:

$$\min_{\Delta q} f(\Delta q)$$

such that: $$J(q)\Delta q = \Delta X$$

$$b_L \leq \Delta q \leq b_U$$

where $f$ is the cost function to be minimized, $J(q)$ is the Jacobian matrix and $b_L, b_U$ are bounds on $\Delta q$. Following this formulation, different criterions have been tested. In this paper, we consider the minimum of absolute work [17] which is defined as the sum of absolute values of the works $W_i$ done by each of the $n$ joint torques:

$$C = \sum_{i=1}^{n} |W_i|$$

This criterion is local and guarantees the minimum energy consumption at each time, but not on the whole trajectory. However, it is biologically plausible and the authors found strong similarities between joint angles obtained from their model and observations in human experiments. Note that the iterative numerical technique used in [17] was limited to a 4 DOF model. In this paper, considering the proposed formulation, it was possible to solve the problem for a 6 DOF arm ($n = 6$) by using the Matlab optimization toolbox. The arm dynamics was used to compute the arm torques $\tau$ from the following model:

$$\tau = M(q)\ddot{q} + N(q, \dot{q})\dot{q} + G(q),$$

where $q$, $\dot{q}$, $\ddot{q}$ are the vectors of angular position, velocity and acceleration of the successive links respectively, $M$ is the inertia matrix, $N$ gathers nonlinear terms such as Coriolis effects and $G$ is the gravitational vector.

III. RESULTS

A. The anthropomorphic model

The body-centered and eye-centered control schemes were both tested on an anthropomorphic model of human body including a 6 DOF arm (3 at the shoulder, 1 at the elbow and 2 at the hand) and a 4 DOF eye-neck kinematic chain (2 at the eye and 2 at the neck). This model follows the structure of the humanoid robot HRP2 [18]. The values of mass and inertia considered for the dynamical model also correspond to this physical system. For both control schemes, a proprioceptive delay of 200 ms and a visual delay of 150 ms were considered, according to measurements in human [19], [14]. Bounds on the angular joints amplitude and velocity were considered. During a time interval $\Delta t = 0.01s$, the joint angle variation cannot exceed $2^\circ$ on the arm (4 DOF) and $0.1^\circ$ on the hand (2 DOF). All these values were taken into account to determine the constraints of the optimization problem. For the alignment module in both schemes, we chose the following strategy: when the hand is not seen, its position is estimated from proprioception; when it enters the visual field, its position is estimated from vision. This strategy is simple and consistent with the fact that vision dominates in motor plan computation [20]. In both the arm and the eye-neck kinematic chain, biases were added to joint angles in order to reproduce the proprioception error in human. The values of biases were chosen between $1^\circ$ and $2.5^\circ$ according to psychophysical experiments [16]. Simulations were performed with Matlab with the dynamic model coded in C. We used an interface that allows to visualize the movement as well the visual perception. In average, a simulation run took about 15s.

B. Trajectories

For each control scheme, numerous hand movements were generated toward target positions distributed in the accessible space. Figures 5 and 6 show two sets of 3D hand trajectories, respectively obtained with the body-centered and the eye-centered schemes. These trajectories start from the same hand position and with the same arm configuration, and end at twenty peripheral targets located on a circle of radius 35cm, in the frontal plane. The corresponding velocity curves are presented in the right part of each figure. A first result is that, in first approximation, both control schemes turn out to provide rather similar hand trajectories. Furthermore, as bio-inspired models were used for the arm controller, these trajectories show important characteristics of human reaching movements, namely: they are quasi-straight with a slight curvature and have single peak and bell-shaped velocity profiles. Another important point is that the curvature of trajectories varies continuously as a function of target position along the support circle, as registered in humans in similar conditions [21].
Fig. 7. Movement of 50 cm in the sagittal plane: angular trajectory (up-left), Cartesian hand coordinates (up-right), hand trajectory (down-left), hand velocity (down-right).

Fig. 8. Same movement as in Fig. 7, with three different hand velocities.

Fig. 7 shows the details of joint variations, hand trajectory and velocity profile during a hand movement of 50cm in the sagittal plane, obtained with the eye-centered scheme. Note that, velocity curves have an asymmetry, which was experimentally observed in human in [22] and [23]. We also validated that the final arm configuration does not depend on the movement speed, as shown in Fig. 8. This result is consistent with experimental observations [22], [24].

C. Effect of biases and sensory delay

Though reaching movements obtained with both control schemes are resembling, a more careful analysis shows that the influence of proprioceptive biases is not equal in both cases. The difference comes from the fact that the difference vector, which constitutes the control input, is not known with the same precision in both schemes. In the eye-centered scheme, this vector is always correctly directed toward the target whereas, due to the proprioception error in the arm and eye-neck joints, the hand position is known with a lower precision. On the other hand, in the body-centered scheme, proprioception through arm joints provides a better localization of the hand with respect to the body but, due to biases in the eye-neck joints, the difference vector does not points correctly to the target. This phenomenon is illustrated in Fig. 9, where the evolution of the difference vector, for each control scheme, is shown during a reaching movement in the frontal plane. The resulting hand trajectories are reported in Fig. 10. Clearly, the trajectory generated in the body-centered frame is more curved than the one generated in the eye-centered frame. This comes from the fact that, the pointing error of the difference vector makes the hand move away from the reference line that links the initial hand position and the target. When the hand enters the visual field, the hand localization error is reduced and the trajectory is bent to the target. It can be also observed in Fig. 5 and 6, that the final part of trajectories obtained with the body-centered scheme is more curved. In our simulation, we also noted that the trajectory deviation is increased by the feedback delay. Although this fact occurs in both reference frames, the increase in eye-centered frame is smaller than in body-centered frame. Here again, it appears that when the difference vector points correctly to the target, the sensibility to sensory delay is reduced, even though the current hand position is known with lower precision.

Fig. 9. Difference vector in eye-centered (top) and body-centered (bottom) frame. The left pink point represents the initial hand position, the right red point represents the actual target position. The vector is blue so long as the the hand is not visible; and red when the it enters the visual field. The jump represents the change of reference direction when the hand becomes visible.

Fig. 10. Trajectories in sagittal plane (Left) and frontal plane (right). Movements in eye-centered frame (blue) and body-centered frame (red).

D. Application to the reaching of a mobile target with HRP2

As an application, both control schemes were tested on the HRP2 humanoid robot of LAAS-CNRS [18]. As the eye and the arm can be simultaneously controlled, an interesting objective was to reach a moving object. Both controllers were successfully applied to this task. Fig. 11 shows an example of movement obtained with the eye-centered scheme. In this task the target started to move two seconds after the onset of the hand movement. The trajectories and velocities of the robot’s
Fig. 11. Reaching movement to a mobile target with HRP2. Hand and object trajectory (left) and velocity (right).

Fig. 12. HRP2 reaching a mobile target with the eye-centered control scheme (Chronological sequence from left to right).

hand and the target are reported in Fig. 12. Note that, due to visual feedback delay, the arm continued its movement a little bit toward the initial target position, while the target had started to move. The images were obtained with the dynamic simulator OpenHRP3. Videos showing different motions are available at http://www.laas.fr/~mtmtnan/works/

IV. CONCLUDING DISCUSSION

Though the proposed control schemes are very far from representing the actual control architecture of voluntary reaching movements in humans, this study provides interesting computational elements towards the complete answer to the three questions posed in the introduction. First, the transformation of the hand-target difference vector with respect to the body does not appear to be a necessary step for controlling the hand movement accurately. As soon as a complete model of the kinematic chain, including the eye-neck joints, is considered, a vision-based control scheme can be used to control precisely the hand movement in a eye-centered frame. Second, our result shows that the transformation of the difference vector with respect to a body-centered frame may induce variations of the hand trajectory. Indeed, though the movement obtained with both control schemes are globally similar, it clearly appears that a wrong pointing direction of the difference vector may steer the movement away from the reference line that links the initial hand position to the target. As a consequence, the global curvature of the trajectory may increase. Third, proprioceptive biases and sensory delays induce a stronger deviation of the trajectory when using the body-centered control scheme. This result proves that the eye-centered scheme is more robust with respect to these perturbations. This observation is consistent with the property of visual servoing in robotics. Indeed, the control is known to be more robust with respect to modeling errors and measurement uncertainties when the reference error is directly expressed in terms of visual features.

ACKNOWLEDGMENT

This work was supported by the Zeuxis project funded by the EADS foundation and by the ROMA project financed by the NeuroInformatique interdisciplinary program of CNRS.

REFERENCES