

Research report

# Temporal prediction of touch instant during observation of human and robot grasping

Laila Craighero<sup>a,\*</sup>, Francesco Bonetti<sup>a</sup>, Luca Massarenti<sup>a</sup>, Rosario Canto<sup>a</sup>,  
Maddalena Fabbri Destro<sup>a</sup>, Luciano Fadiga<sup>a,b</sup>

<sup>a</sup> Department of Biomedical Sciences and Advanced Therapies, Section of Human Physiology,  
University of Ferrara, via Fossato di Mortara 17/19, 44100 Ferrara, Italy

<sup>b</sup> Italian Institute of Technology, Genova, Italy

Available online 12 February 2008

## Abstract

The aim of the present work was to test the ability to predict the instant at which a grasping hand touches an object. Our hypothesis was that, because of the activation of the mirror-neuron system, the same predictive process necessary for action execution should be active during observation. Experimental evidence indicates, however, that not only observed actions but also observed objects automatically activate observer's motor repertoire. What happens, therefore, if the observed action is different from the one automatically evoked by the vision of the object? To answer this question we presented subjects with two different grasping actions: the one most suitable for the presented object and a less appropriate one. Subjects were required to detect the instant at which the demonstrator's hand touched the object. In a further condition, subjects were required to detect the outcome of an action performed by a robotic arm moving with constant kinematics. Results showed that while in the case of robot grasping subjects responded before the touch instant, in the case of human grasping the response followed the touch instant, but occurred much earlier than simple reaction times. This demonstrates that subjects were able to predict the outcome of the seen action. The predictive capability was specifically enhanced during observation of the "suitable" grasping. We interpret these results as an indication of the synergic contribution of both object-related (canonical) and action-related (mirror) neurons during observation of actions directed towards graspable objects.

© 2008 Elsevier Inc. All rights reserved.

**Keywords:** Action observation; Motor resonance; Mirror system; Canonical neurons; Action prediction; Monocular and binocular vision

## 1. Introduction

Since our discovery of mirror neurons we suggested that they might play a role in action understanding [6,9, see 21]. The core of the proposal was the following: when an individual acts she selects an action whose motor consequences are known to her. Mirror neurons allow this knowledge to be extended to actions performed by others. Each time an individual observes an action done by another individual, neurons that represent that action are activated in her premotor cortex. Because the evoked motor representation corresponds to that internally generated during action execution, the observer "understands" the other's action [see 23].

In favor of this idea are the results of a single neuron recording study performed in monkeys by Umiltà et al. [29]. The experi-

mental paradigm consisted of two basic conditions. In one, the monkey was shown with a fully visible action directed towards an object. In the other, the monkey saw the same action but with its final critical part hidden by a curtain. Before each trial the experimenter placed or not a piece of food behind the screen so that the monkey knew whether or not a target for the action was there. The main result of the experiment was that several tested neurons discharged in the 'hidden' condition, but only when the animal knew that the food was present. This experimental evidence was interpreted as a good demonstration that mirror neurons fire also when the reaching/grasping of an object occurs out of sight, as long as the intention and the plausibility of the action are clear. This because the understanding of the action is not only based on the visual description of the scene but it relies also on the motor representation of the action goal, shared by both the agent and the observer, and triggered by the context in which the action is performed (i.e. the presence or the absence of the food on the table behind the screen). These results strongly suggest that the mirror-neuron system is more

\* Corresponding author.

E-mail address: crh@unife.it (L. Craighero).

an “active interpreter” than a “passive perceiver” [see also 7]. Thus, the involvement of the observer’s motor system is not only evoked by the visual features of the observed action but, rather, it may intervene in filling gaps because it gives the observer an implicit motor knowledge of what other individuals are doing. In other words, the mirror system should possess the capability to predict the action outcome. In a recent experiment, Kilner et al. [14], by using event-related potentials in humans, showed that the readiness potential (*Bereitschaftspotential*), a well known electrophysiological marker of motor preparation, is detectable also during action observation. In the same direction point the data by Borroni et al. [1] showing that the corticospinal pathway of the observer is modulated according to the temporal pattern of muscle recruitment in the agent and not to the temporal development of the seen movement. In fact, they demonstrated that, during observation of cyclic wrist flexion/extension, the motor facilitation induced in the observer (H-reflex) was temporally congruent with agent’s muscles recruitment, which naturally occurred well before the visible movement.

The hypothesis at the basis of the present work was that to visually detect the instant at which a grasping hand touches an object, subjects should implicitly use the internal model of the seen action. However, objects *per se* possess the capability to automatically activate the motor program more adapt to interact with them. This has been shown at the single neuron level in monkeys [16] and by psychophysics [5,28] and brain imaging [10] experiments in humans. To explore these issues, we tested if the internal motor representation of the observed action is at the basis of the capability to predict the instant at which the demonstrator’s finger pads touch the target object during grasping. Moreover, we manipulated the degree of suitability of the observed grasping action by presenting the subjects with two different grasping movements: the one most suitable for the presented object (see Fig. 1b, sagittal grasping) and a less appropriate one, orthogonal to the first (see Fig. 1a, frontal grasping). During a grasp, the final finger position defines an opposition axis through which opposite forces operate on the object [17,11]. The orientation of this axis is constrained by the biomechanics of the arm, which determine more or less comfortable hand postures [8]. In the present experiment the more comfortable hand posture is represented by the sagittal grasping which requires less forearm rotation before grasping, in contrast with the frontal grasping which imposes a hand supination

motion. The two different grasping movements give origin to two situations very different also in terms of visual perception. Indeed, while in the frontal grasping the two fingers are both visible and act at approximately the same distance from subject’s frontal plane, in the sagittal grasping the thumb is hidden by the object and the two fingers act at different depth. Therefore, in order to investigate the role of stereopsis during detection of the touch instant, we tested subjects during both monocular and binocular vision. If the detection of the time-to-contact is exclusively based on visual information and does not involve the motor system of the observer, it is possible that monocular vision, which prevents a fine in-depth detection, would worsen subject’s responses, particularly during observation of the sagittal grasping, which develops in depth, perpendicularly to the subject’s frontal plane.

Finally, to have a control condition in which no internal model is necessary to detect the time-to-contact since almost no variability is present in the observed event, we asked subjects to detect the instant at which a robotic arm touched the target object during grasping. The robot always moved with the same kinematics generating a predictable movement, time-locked to the instant at which its arm started to move.

## 2. Materials and methods

### 2.1. Subjects

Eleven undergraduate students of the University of Ferrara participated in the experiment after giving their informed consent. All subjects were right-handed according to the Oldfield norms [18]. Five subjects were left ocular dominant and six were right ocular dominant, as assessed by a series of classical tests of sighting dominance (see [20]) such as: (i) questions on which eye is the favorite one to look into a hole. (ii) The “hole-in-the-card test” during which the subject is requested to look at a distant object with both eyes open through a hole made on a card. The observer then alternatively closes each eye to determine which one is viewing the object (i.e. the dominant eye); (iii) the observer extends one arm, then with both eyes open aligns the index finger with a distant object. The observer then alternatively closes each eye to determine which one is viewing the object.

### 2.2. Procedure

Subjects were seated on a comfortable chair with the head positioned on a chin holder. They were instructed to “tap the desk with your right index finger at the instant at which the experimenter will touch the to-be-grasped object”. Subjects’ responding hand was kept out of sight. The pad of subjects’ right index

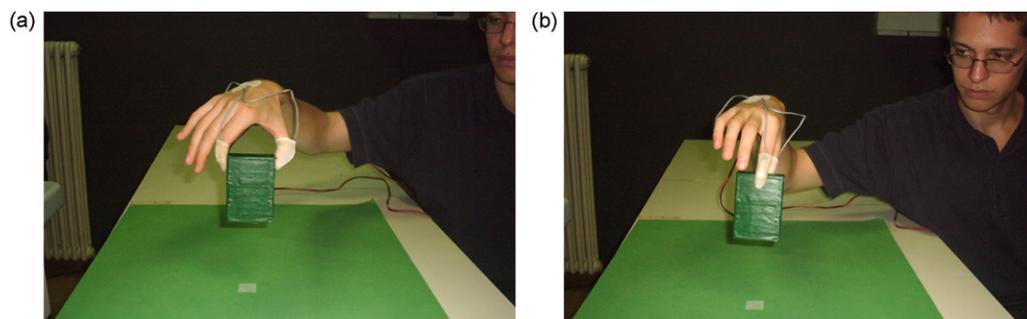


Fig. 1. Example of frontal (left panel) and sagittal (right panel) grasping movements as performed by the demonstrator in front of the subject. Note the touch sensors attached to both the index finger and the thumb of the demonstrator.

finger as well as the pads of experimenter's index finger and thumb were covered by a resistive touch sensor (strain gage). The capability to predict action's outcome was measured by the time lag between the instant at which the demonstrator touched the object with either index finger or thumb and subject's response time. The time lag was positive when the subject's response was subsequent to the touch of the object, and negative when the subject's response anticipated the touch. Subjects' and experimenter's sensors were connected to Wheatstone bridges, whose outputs were amplified, digitized by an A/D converter and sent to a personal computer for the successive analysis. As "touch instant" was considered the instant at which a 1% increase of sensor's signal with respect to the baseline was detected by the analysis software.

The experimenter, seated in front of the subject, grasped with a natural velocity precision grip, a plastic parallelepiped (9 cm × 6 cm × 2.5 cm, see Fig. 1) located 70 cm from subject's frontal plane. Two kinds of grasping were performed: in the first ("not suitable grasping"), the fingers' opposition space was parallel to subject's frontal plane (i.e. index finger and thumb both visible, Fig. 1, left panel). In the second ("suitable grasping"), the thumb was occluded by the object, while the index finger was visible. Note that in this latter case, finger closure took place along subject's sagittal plane (Fig. 1, right panel). The two different grasping movements had similar duration (~900 ms). To control this, we video recorded one single experimental session and, by using a specific software (Adobe Premiere 1.5), we calculated the movement time of the two grasping movements. The suitability of the grasping was tested in a preliminary experiment involving 12 subjects different from those participating to the experiment, and demonstrating that participants almost constantly selected (88% of observations) the "suitable grasping" when requested to grasp the same object used in our experiment.

A third type of grasping, "robot grasping", was presented to the subjects, consisting in the movement of a robotic arm towards the same object grasped by the experimenter. We used here the same robotic arm used by Castiello et al. [4]. This robotic arm, kindly provided us by Castiello, looked like an average human forearm with a gloved hand and used a single motor to move from a vertical to a horizontal position (1 degree of freedom). The construction was electro-mechanical and controlled by an 87c751 micro-controller. The four fingers and the thumb had a common movement, so as to mimic the closing of a human hand. The hand construction was nylon cords for tendons, silicon rubber for joints, and wooden dowels for bones. The movement was provided by a dc electric motor that tenses the tendons to close the hand. Limit sensors on the arm and hand were used by the micro-controller to control movement. The arm length was approximately 0.5 m. The movement of the robot was quite smooth and the folding of the hand was comparable to a 'human' grasping action. The robot was programmed to simultaneously move its arm and open its fingers when the experimenter pressed a button. The maximum aperture of the fingers (110 mm) was programmed to occur after ~72% of the total movement duration (800 ms) had elapsed. This value corresponds to the occurrence of maximum grip aperture usually found in adult humans when performing a whole hand prehension [2]. After reaching the maximum aperture the fingers started to close upon the to-be-grasped object. Thus, the temporal and kinematics parameters of the movement were always the same in all trials.

Since the movement of the robotic arm was quite noisy, to avoid the possibility to detect the time-to-contact exclusively on the basis of acoustical cues, the subjects wore ear-phones playing music during all experimental conditions.

Subjects were required to observe not suitable (frontal, 15 trials), suitable (sagittal, 15 trials) and robot graspings (30 trials) randomly presented during three different experimental sessions differing for the type of vision: binocular, monocular dominant and monocular non-dominant. The order of these sessions was randomized across subjects. In the monocular sessions, a patch was used to cover the eye not involved in the task.

Data were analyzed online and demonstrator's touch and subject's response were recorded on the computer hard disk for the successive statistical analysis (see below).

### 3. Results

Touch sensors data (i.e. the difference in time between demonstrator's touch and subject's response) were submitted off-line to an analysis of variance. Type of grasping (not suitable,

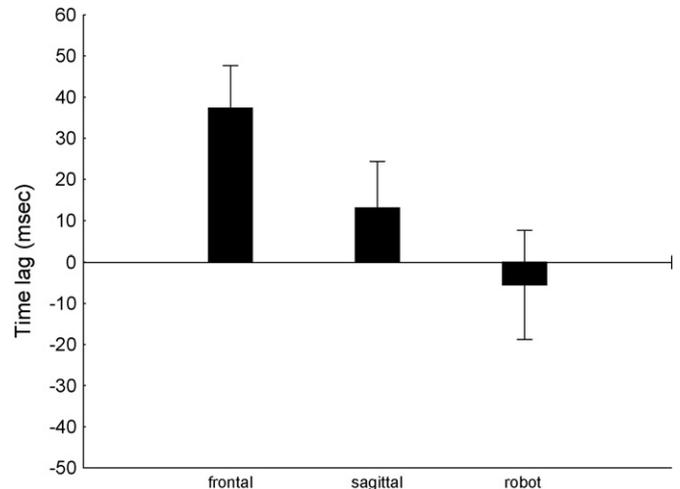


Fig. 2. Time lag between the instant at which the demonstrator touches the object and subject's response time. Data for frontal, sagittal, and robot grasping movements are shown. Thin lines above histograms indicate standard error of mean. Ordinates are in milliseconds.

suitable, and robot) and type of vision (binocular, monocular dominant, monocular non-dominant) were the within-subject factors. The analysis of variance showed that both main factors, type of grasping and type of vision, were statistically significant ( $F(2,20) = 6.13, p < 0.01$  and  $F(2,20) = 3.62, p < 0.05$ , respectively). The interaction between the two factors was not significant.

Data shown in Figs. 2 and 3 suggest that two main results arise from the present study.

The first result is that, during hand action observation, the time lag between the instant at which the demonstrator touches the object and that of subjects' response depends on the suitability of the observed grasping (Fig. 2): while not suitable grasping

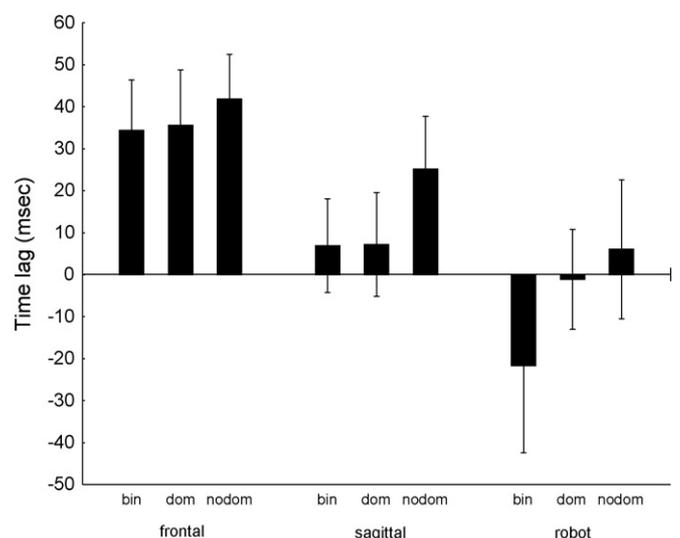


Fig. 3. Time lag between the instant at which the demonstrator touches the object and subject's response time. Data for binocular (bin), monocular dominant (dom) and monocular non-dominant (nodom) conditions are shown for each type of observed movement (frontal, sagittal, and robot grasping). Thin lines above histograms indicate standard error of mean. Ordinates are in milliseconds.

is detected with an average delay of 37 ms, in the case of suitable grasping the time lag is 13 ms ( $T$ -test,  $p < 0.01$ ).

However, both values indicate that subjects are indeed predicting (and not reacting at) the instant at which the demonstrator touches the object, since the time lags are definitely shorter than the fastest simple reaction time. Finally, when subjects are requested to detect the touch of robot grasping, the time lag is further reduced and becomes negative ( $-5$  ms), that is subject's response precedes the instant at which the demonstrator's finger touches the object, mainly because of the contribution of the binocular vision condition (Fig. 3). Indeed, although not significant, the trend shown in Fig. 3 in the robot condition (binocular < monocular dominant < monocular non-dominant) suggests that while observing a robot-arm, we base our prediction mainly on visual cues.

The second result of this study is that, during observation of human grasping, predictive capability does not depend upon stereoscopic vision. Fig. 3 shows that the average time lags for both frontal and sagittal grasping type are essentially the same for binocular and monocular vision. The increase present in monocular non-dominant vision is not significant.

#### 4. Discussion

In the present work we wanted to verify if the suitability of the observed grasping modulates the ability to predict the touch instant. To this purpose two different types of grasping were presented, differing for the type of fingers opposition space: in one case the type of grasping was the one more commonly chosen to grasp the presented object, in the other case it was a less appropriate one. By showing to the subjects two type of grasping on the same object, i.e. the suitable and the non-suitable one, we voluntarily created a situation of conflict in terms of motor representations. Accordingly, two main conditions characterize the present experiment: a congruent one, in which the motor program evoked by object observation coincides with that executed by the experimenter, and an incongruent one, where the two motor programs differ. The presence of a difference in the performance according to the suitability of the observed grasping would therefore reflect the congruence between the motor representation evoked by the observation of the object and that evoked by the observation of the action.

Our results show that subjects' response times are well below those commonly found in simple reaction times tasks (usually around 120–150 ms), indicating that, to accomplish the task, subjects indeed use a predictive model of the seen action. Moreover, response times were shorter for suitable grasping trials than for not suitable ones. This indicates that action prediction is based on the internal motor representation of the seen action, and that whenever an incongruence is present between the action evoked in the observer by the to-be-grasped object and the observed action, actually executed on it, the ability to predict the action outcome decreases.

It is important to note that this effect is independent from the type of vision (binocular, monocular), therefore excluding

the possibility that the detection of the time-to-contact is exclusively based on visual information. In fact, since the sagittal, suitable, grasping occurs mainly along subjects' sagittal plane it should have benefited more from depth perception than the frontal one. Results show that this was not the case. In a very recent experiment Pozzo et al. [19] reached a similar conclusion by testing whether motion inference relies only upon visual extrapolating mechanisms based on past visual trajectory information. They asked subjects to estimate the vanishing or final positions of a dot moving either upwards or downwards that was masked in the last part of the trajectory. The stimulus could either move according to biological or non-biological kinematic laws. They found that precision in estimation was greater when stimulus' kinematics belong to subjects' motor repertoire.

The significance of the factor type of vision, however, indicates that, independently from the type of grasping to be detected, binocular vision allows a better prediction of the touch instant. This last result is consistent with findings indicating that removing binocular vision does affect prehension in terms of kinematics [12,15,25,26]. Servos [24] interprets this effect not as a consequence of a purely perceptual bias but rather as a visuo-motor effect, due to a monocular input to a system that normally calibrates motor output on the basis of binocular vision.

Finally, as far as the robot grasping is concerned, the present experiment shows a significant difference between observation of actions performed by the experimenter and those performed by the robotic arm: in the latter case the time lag is further reduced and becomes negative. This last result fully agrees with that shown by Wohlschlagel et al. [30], showing that the instant of button press executed by a robot arm is identified in advance to its actual occurrence. Electrophysiological data on monkeys [9,22], and brain imaging [13,27] and psychophysical studies on humans [3,30], indicate that the mirror-neuron system resonates only during observation of biological actions. Thus, our hypothesis was that the prediction of the robotic arm's action outcome would not be based on the internal motor representation. Moreover, while the temporal and kinematics parameters of robot movements are constant trial by trial, human grasping, even if replicating exactly the same action, is variable by definition. As a consequence, the prediction of the robot's action outcome can be merely achieved by visually detecting its starting time, without an involvement of the observers' motor representations.

In conclusion, the present study indicates that, in humans, the capability to predict the instant at which an observed grasping hand touches an object is indeed possible, and that this capability depends on the congruence between the observed action and the internal motor representation activated in the observer by the vision of the object.

#### Acknowledgements

We thank Umberto Castiello for allowing us to use the robotic arm. This work has been supported by European Commission grants ROBOT-CUB, CONTACT and NEUROBOTICS to L.F. and L.C., and by Italian Ministry of Education grants to L.F.

## References

- [1] P. Borroni, M. Montagna, G. Cerri, F. Baldissera, Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement, *Brain Res.* 1065 (2005) 115–124.
- [2] U. Castiello, Grasping a fruit: selection for action, *J. Exp. Psychol. Hum. Percept. Perform.* 22 (1996) 582–603.
- [3] U. Castiello, Understanding other people's actions: intention and attention, *J. Exp. Psychol. Hum. Percept. Perform.* 29 (2003) 416–430.
- [4] U. Castiello, D. Lusher, M. Mari, M. Edwards, G.W. Humphreys, Observing a human or a robotic hand grasping an object: differential motor priming effects, in: W. Prinz, B. Hommel (Eds.), *Attention and Performance: vol. XIX. Common Mechanisms in Perception and Action*, Oxford University Press, Oxford, England, 2002, pp. 315–333.
- [5] L. Craighero, L. Fadiga, C.A. Umiltà, G. Rizzolatti, Evidence for visuo-motor priming effect, *Neuroreport* 8 (1996) 347–349.
- [6] G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, G. Rizzolatti, Understanding motor events: a neurophysiological study, *Exp. Brain Res.* 91 (1992) 176–180.
- [7] L. Fadiga, L. Craighero, M. Fabbri-Destro, L. Finos, N. Cotillon-Williams, A.T. Smith, Language in shadow, *Soc. Neurosci.* 1 (2006) 77–89.
- [8] V. Frak, Y. Paulignan, M. Jeannerod, Orientation of the opposition axis in mentally simulated grasping, *Exp. Brain Res.* 136 (2001) 120–127.
- [9] V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, Action recognition in the premotor cortex, *Brain* 119 (1996) 593–609.
- [10] J. Grèzes, M. Tucker, J. Armony, R. Ellis, R.E. Passingham, Objects automatically potentiate action: an fMRI study of implicit processing, *Eur. J. Neurosci.* 17 (2003) 2735–2740.
- [11] T. Iberall, G. Bingham, M.A. Arbib, Opposition space as a structuring concept for the analysis of skilled hand movements, *Exp. Brain Res. (Suppl)* 15 (1986) 158–173.
- [12] S.R. Jackson, C.A. Jones, R. Newport, C. Pritchard, A kinematic analysis of goal-directed prehension movements executed under binocular, monocular and memory-guided viewing conditions, *Vis. Cogn.* 4 (1997) 113–142.
- [13] J. Jarvelainen, M. Schurmann, S. Avikainen, R. Hari, Stronger reactivity of the human primary motor cortex during observation of live rather than video motor acts, *Neuroreport* 12 (2001) 3493–3495.
- [14] J.M. Kilner, C. Vargas, S. Duval, S.J. Blakemore, A. Sirigu, Motor activation prior to observation of a predicted movement, *Nat. Neurosci.* 7 (2004) 1299–1301.
- [15] A. Loftus, P. Servos, M.A. Goodale, N. Mendarozqueta, M. Mon-Williams, When two eyes are better than one in prehension: monocular viewing and end-point variance, *Exp. Brain Res.* 158 (2004) 317–327.
- [16] A. Murata, L. Fadiga, L. Fogassi, V. Gallese, V. Raos, G. Rizzolatti, Object representation in the ventral premotor cortex (area F5) of the monkey, *J. Neurophysiol.* 78 (1997) 2226–2230.
- [17] J.R. Napier, The prehensile movements of the human hand, *J. Bone Joint Surg.* 38B (1956) 902–913.
- [18] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, *Neuropsychologia* 9 (1971) 97–113.
- [19] T. Pozzo, C. Papaxanthis, J.L. Petit, N. Schweighofer, N. Stucchi, Kinematic features of movement tunes perception and action coupling, *Behav. Brain Res.* 169 (2006) 75–82.
- [20] R.W. Reading, *Binocular Vision—Foundations and Applications*, Butterworths, Boston, 1983.
- [21] G. Rizzolatti, L. Craighero, The mirror-neuron system, *Annu. Rev. Neurosci.* 27 (2004) 169–192.
- [22] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi, Premotor cortex and the recognition of motor actions, *Cogn. Brain Res.* 3 (1996) 131–141.
- [23] G. Rizzolatti, L. Fogassi, V. Gallese, Neurophysiological mechanisms underlying the understanding and imitation of action, *Nat. Rev. Neurosci.* 2 (2001) 661–670.
- [24] P. Servos, Distance estimation in the visual and visuomotor systems, *Exp. Brain Res.* 130 (2000) 35–47.
- [25] P. Servos, M.A. Goodale, Binocular vision and the on-line control of human prehension, *Exp. Brain Res.* 54 (1994) 121–132.
- [26] P. Servos, M.A. Goodale, L.S. Jakobson, The role of binocular vision in prehension: a kinematic analysis, *Vis. Res.* 32 (1992) 1513–1521.
- [27] Y.F. Tai, C. Scherfler, D.J. Brooks, N. Sawamoto, U. Castiello, The human premotor cortex is 'mirror' only for biological actions, *Curr. Biol.* 14 (2004) 117–120.
- [28] M. Tucker, R. Ellis, On the relations between seen objects and components of potential actions, *J. Exp. Psychol. Hum. Percept. Perform.* 24 (1998) 830–846.
- [29] M.A. Umiltà, E. Kohler, V. Gallese, L. Fogassi, L. Fadiga, C. Keysers, G. Rizzolatti, I know what you are doing. A neurophysiological study, *Neuron* 31 (2001) 155–165.
- [30] P. Wohlschlagel, P. Haggard, B. Gesierich, W. Prinz, The perceived onset time of self- and other-generated actions, *Psychol. Sci.* 14 (2003) 586–591.