Biol Cybern (2010) 103:471–485
DOI 10.1007/s00422-010-0415-5

ORIGINAL PAPER

Biological Cybernetics

From motor to sensory processing in mirror neuron computational modelling

Giovanni Tessitore · Roberto Prevete · Ezio Catanzariti · Guglielmo Tamburrini

Received: 17 March 2010 / Accepted: 29 November 2010 / Published online: 17 December 2010

Abstract Typical patterns of hand-joint covariation arising in the context of grasping actions enable one to provide simplified descriptions of these actions in terms of small sets of hand-joint parameters. The computational model of mirror mechanisms introduced here hypothesizes that mirror neurons are crucially involved in coding and making this simplified motor information available for both action recognition and control processes. In particular, grasping action recognition processes are modeled in terms of a visuo-motor loop enabling one to make iterated use of mirror-coded motor information. In simulation experiments concerning the classification of reach-to-grasp actions, mirror-coded information was found to simplify the processing of visual inputs and to improve action recognition results with respect to recognition procedures that are solely based on visual processing. The visuo-motor loop involved in action recognition is a distinctive feature of this model which is coherent with the direct matching hypothesis. Moreover, the visuo-motor loop sets the model introduced here apart from those computational models that identify mirror neuron activity in action observation with the final outcome of computational processes unidirectionally flowing from sensory (and usually visual) to motor systems.

G. Tessitore (✉) · R. Prevete · E. Catanzariti · G. Tamburrini
Dipartimento di Scienze Fisiche, Università di Napoli Federico II,
Via Cintia Monte S. Angelo, 80126 Naples, Italy
e-mail: tessitore@na.infn.it

R. Prevete
e-mail: prevete@na.infn.it

E. Catanzariti
e-mail: catanzariti@na.infn.it

G. Tamburrini
e-mail: tamburrini@na.infn.it

Keywords Mirror neurons · Direct-matching hypothesis · Motor knowledge · Motor-driven sensory processing · Grasping actions

1 Introduction

Mirror neurons exhibit the special behavioral property of becoming active during both execution and observation of object-directed actions. Identified in the macaque's F5 cortical motor area, these neurons were first described in seminal work by Rizzolatti et al. (1996; Gallese et al. 1996). According to a prominent interpretation, mirror neurons are involved in a circuit of cortical areas—usually referred to as mirror system (Cattaneo and Rizzolatti 2009; Rizzolatti and Craighero 2004)—subserving the control of one's own actions and the recognition of observed actions. This interpretation posits significant functional commonalities between action control and action recognition processes, which take their origin in a set of shared neurobiological mechanisms. Additional functional roles have been hypothesized for mirror systems in the framework of theories of language evolution (Arbib 2005), mind-reading (Gallese and Goldman 1998), and imitation learning by humans (Carr et al. 2003; Miall 2003).

Various computational models have been advanced to account for mirror neuron functional roles in the broader context of mirror system functionalities (see Oztop et al. 2006 for a relatively recent review). Typically, these models construe mirror neuron behaviors as the outcome of computational processes unidirectionally flowing from sensory (and mostly visual) input to mirror neuron activity. According to these models, substantive computational work is carried out by the visual system, and mirror neuron behavior occurs as a side effect of view-independent visual classification abilities. A tension arises between these computational accounts and the *direct matching hypothesis* (Rizzolatti et al. 2001), which assigns a more central role in action recognition processes to the motor system in general, and to mirror neurons located in macaque F5 brain area in particular.

A different approach to the computational modelling of mirror neurons, inspired by the direct matching hypothesis, is pursued here. Indeed, mirror activity is hypothesized to code for motor information which is supplied to the visual system for the purpose of interpreting sensory inputs. More specifically, this model assigns a central functional role to mirror mechanisms in visual perception, predicting that motor information coded by mirror neurons simplifies the processing of sensory (visual) inputs and improves the results of action recognition tasks. Moreover, the motor information this model attributes to the mirror system during the control of one's own actions is such that an inhibition of mirror neuron activity results in action performance slowdown rather than overall action performance suppression.

Central biological findings supporting this computational modelling approach can be roughly summarized as follows: (a) mirror neurons are mainly located in a brain motor area which subserves the control of object-directed actions; (b) when a monkey performs an object-directed action, an induced inactivation of F5 mirror neurons slows down motor activity without suppressing it, since the monkey is eventually able to perform the correct action (Fogassi et al. 2001); (c) mirror neuron activity encodes only a fragment of the total motor information which is stored and deployed in area F5 to execute similar actions.

The conjunction of (a) and (b) suggests that information carried by mirror neurons facilitates action control without being strictly necessary to complete action execution. And mirror activation in observation conditions suggests, in the light of (c) that motor information stored in F5 is involved, but selectively so, in perceptual action recognition processes.

Within this broad theoretical framework, the following questions are specifically addressed: *What* is the information coded by mirror neuron activity? *How* does the information coded by mirror neurons bear on sensory processing?

The answer to the *what* question builds up on the suggestion, that hand-joint configurations can be expressed by a small set of parameters (Mason et al. 2001), insofar as typical patterns of hand-joint covariation arise in the context of grasping actions. More specifically, we propose that the set of object-directed actions can be subdivided into distinct classes, each one of which is identified by means of a small set of vectors spanning a subspace in the space of hand-joints configurations. Thus, each hand-configuration can be expressed by suitably setting coefficients in a linear combination of these vectors. These subspaces will be called *action subspaces* here (see Sect. 3.1). Accordingly, if one knows which action subspace is associated to some given class of object-directed actions, then a simplified control of each action belonging to that class can be achieved by setting a relatively small number of parameters. The functional role of mirror neuron activity during action execution is identified here with a crucial component of the action subspace selection process.

The answer to the "how" question urges one to revise a central hypothesis underlying most computational models of mirror neurons and systems. According to this hypothesis, action recognition is based on the mapping from sensory (and mostly visual) input to a view-independent action representation, which is expressed in terms of large numbers of kinematic parameters including hand-joint configuration and wrist velocity.

In contrast with this, the task of transforming sensory input into intrinsic action features is simplified on the basis of motor information expressed in terms of action subspaces. In particular, sensory input is mapped on each grasping action subspace codified in the motor system using a set of specialized submappings. And the more likely submapping is selected by mirror neuron activity.

On the whole, the computational model introduced here assigns mirror neurons a functional role in appropriate action subspace selection, which is crucially involved in both action execution and action observation processes.

In Sect. 2, distinguishing features of this computational model are emphasized by a comparison with major extant computational models of mirror neurons. In Sect. 3, the "what" question concerning motor information coded by mirror neurons during both action execution and observation is addressed in a more detailed fashion. And a computational account of the interaction between motor information and perceptual (visual) processes is provided within a probabilistic framework. In Sects. 4 and 5, model behaviors observed in simulation experiments that are carried out in the context of reach-to-grasp actions are described and analyzed. These results show that visual sensory processing benefit in various experimental settings from the interaction with motor knowledge.

2 Unidirectional visuo-motor transformations

According to the direct matching hypothesis, the motor system plays a central role in action recognition. An action is recognized by an observer when its observation brings a part of the observer's neural motor system to resonate. More specifically, "the 'motor knowledge' of the observer is used to understand the observed action. In other words, we understand an action because the motor representation of that action is activated in our brain." (Rizzolatti et al. 2001, p. 661) Even though this hypothesis provides a convincing functional interpretation of experimental data, it is still unsatisfactory from a computational point of view,

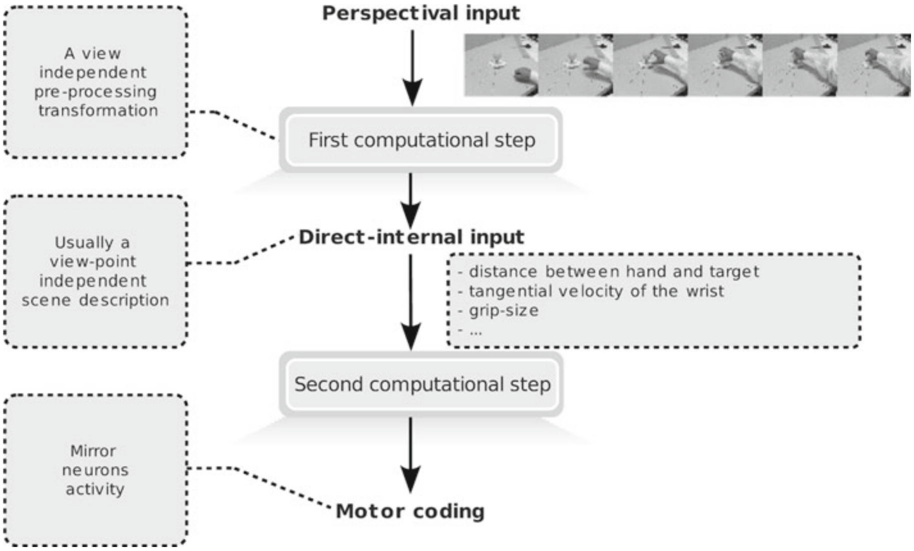

Perspectival input
A view independent pre-processing transformation
First computational step
Usually a view-point independent scene description
Direct-internal input
- distance between hand and target
- tangential velocity of the wrist
- grip-size
- ...
Second computational step
Mirror neurons activity
Motor coding

Fig. 1 Most computational models account for mirror neuron activity in terms of a unidirectional flow of computation, initially turning perspectival sensory input into direct internal input, and then transforming direct internal input into motor coding. Accordingly, the motor system is the target of computations taking place elsewhere, and plays hardly any substantive functional role in action recognition

insofar as it does not specify what is the motor knowledge coded by mirror neurons and how it is used for action recognition purposes. Let us now turn to address these "what" and "how" issues.

To begin with, let us notice that one can envisage different sorts of involvement for motor representations and processing in action recognition. Notably, the following definitions are conducive to distinguishing between two major views of how the motor system is involved in action recognition:

- *Direct internal input* is from now on identified with the complex of brain signals that mirror neurons and F5 neurons closely involved in mirror neuron activity receive from directly afferent brain areas during observation or execution of object-directed actions.
- *Perspectival sensory input* is from now on identified with the complex of sensory (visual) and proprioceptive inputs collected from the perspective of observer or executor of object-directed actions.

According to one view of how the motor system becomes active during action recognition processes, perspectival sensory input is turned into motor system activation by a cascade of two computational steps: (1) a transformation unidirectionally flowing from perspectival sensory input to direct internal input, and (2) a transformation unidirectionally flowing from direct internal input to motor coding (see Fig. 1).

This view of motor system involvement in action recognition is endorsed, as we shall presently see, by prominent computational models of mirror neurons. An additional assumption underlying most computational models of mirror neurons is that one computes the same direct internal input in action recognition mode which is computed in action execution mode. In other words, for each object-directed action $\mathcal{A}$, the outcome of the first computational step is the *same* direct internal input, irrespectively of whether $\mathcal{A}$ is observed or executed. This schematic account of sensory-to-motor system mapping expresses a strong modelling commitment, insofar as perspectival inputs of action executors do not, in general, coincide with perspectival inputs of action observers. Consider, for example, visual sensory inputs of some object-directed action $\mathcal{A}$ collected from the perspective of the executor (see Fig. 2) and of the observer (see Fig. 3). Thus, substantive pre-processing of sensory inputs is in many cases required to converge on the same direct internal input from these different perspectives (see Prevete et al. 2008; Fleischer et al. 2009, for a precise computational analysis of view-independent mappings).

A computational model positing unidirectional visual-to-motor transformations is the MNS1 model (Oztop and Arbib 2002). This model embodies the so-called hand-state hypothesis, according to which one computes a hand-state vector, thereby making available view-independent information about hand/target-object pairs. In the MNS1 model, sequences of hand-state vectors are the input data for an action classification module, whose output is identified with

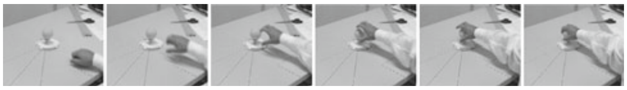


Fig. 2 Visual sensory input to an executor of some object-directed action $\mathcal{A}$

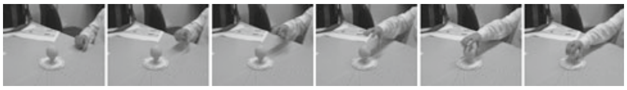


Fig. 3 Visual sensory input to an observer of some object-directed action $\mathcal{A}$

mirror neuron activity in both observation and execution modes. Therefore, this model of mirror mechanisms presupposes substantive pre-processing steps, which enable one to extract view-independent information about object-directed actions from both executor and observer sensory inputs. Likewise, in Haruno et al. (2001), Keysers and Perrett (2004), Ito and Tani (2004), Oztop et al. (2005), and Oztop and Arbib (2002), one presupposes that a view-independent description of the environment is computed and fed into mirror neuron mechanisms.

A different view of how the motor system is involved in action recognition is suggested by the direct-matching hypothesis. The computational approach presented in the next section, inspired by the direct matching hypothesis, is characteristically shaped by the following modelling commitments:

- *In action execution mode* a specific functional interpretation is provided for mirror neuron activity, which turns out to facilitate action control without being strictly necessary to performing the action.
- *In action observation mode* (i) mirror neuron activity is involved at earlier stages of sensory input processing; (ii) motor information coded by mirror neuron activity and identified in action control processes (Rizzolatti and Sinigaglia 2010) is used in observation mode for the purpose of interpreting perspectival sensory inputs.

3 Mirror neurons in sensory-motor loops

To begin with, let us state the main assumptions underlying the computational modelling approach proposed here.

1. *Object/context information* a necessary condition for F5 mirror neuron activity to arise in observation modality is the evidence (usually perceptual evidence) that there is an object towards which the action is directed. This assumption is strongly supported by neurophysiological findings (Rizzolatti et al. 1996; Umilta' et al. 2001).
2. *Affordance perception* given that the observer knows how to manipulate the object, then the observer knows the set of the more likely configurations a hand (more generally, an effector) can sequentially assume in actions that are directed towards that object. This assumption is some sort of specialization to hand configurations of the affordance perception hypothesis in the psychology of perception (Gibson 1979), according to which an animal selectively identifies opportunities for interacting with objects in the environment.
3. *Motor knowledge in perception* knowledge of hand (effector) configurations, which is codified in motor areas, can be used to form a priori hypotheses which constrain the computation of a mapping from perspectival sensory input to hand (effector) configuration coding.
4. *Perceptual-motor loop* using the outcomes of this mapping jointly with perspectival sensory inputs, one can iteratively corroborate or replace previously advanced a priori hypotheses.

The meaning and evidence for assumptions 3 and 4 will be more extensively examined in the next section, starting from the question “What kind of motor information can prove useful in both action control and recognition?”

3.1 Hand eigenpostures and sensory processing

It is well known that suitable forms of motor knowledge can reduce the complexity of control problems. In the context of grasping control, for example, the opportunity of relying on simplified hand description models has been explored on these grounds. In particular, in Iberall and Fagg (1996) and Iberall et al. (1986), the notion of virtual finger is introduced in order to reduce the degrees of freedom one has to take into account. And a simplified control strategy for reach-to-grasp actions is suggested, at the output stage, by the restricted number of hand shapes that qualify as effective grasp configurations. Coordinated movements of hand fingers result, during reach-to-grasp actions, into a reduced number of physically possible hand shapes (see Santello et al. 2002; Mason et al. 2001). Notably, in Mason et al. (2001) a principal component analysis (PCA) (Hastie et al. 2003) is performed over a series of hand-joints configurations that are monitored while a subject performs (or mimics) reach-to-grasp actions.

It is well know that PCA is a linear technique for dimensionality reduction. The output of PCA analysis is a set of orthogonal vectors (called principal components) which enable one to express input data as a linear combination of principal components. Each principal component is associated with a scalar value (called eigenvalue) which expresses input data variance along the associated component. Principal components are sorted on the basis of the associated eigenvalues in decreasing order. Thus, the first principal component is the direction in feature space along which projections have the largest variance. The second principal component is the direction which maximizes variance among all directions that are orthogonal to the first component, and so on.

Formally, let H be an $n \times c$ matrix containing the input vectors row-wise arranged. In this case, each input vector is identified with a hand-joint configuration. Then, principal components are computed by means of the following basic steps:

- Compute the average vector $\bar{\mathbf{hc}} \equiv (\bar{hc}_1, \ldots, \bar{hc}_c)$ as: $\bar{hc}_j = \frac{1}{n}\sum_{i=1}^{n} H_{ij}$ for $j = 1, \ldots, c$;

- Compute $\bar{H}$ by subtracting $\bar{\mathbf{hc}}$ to each row in H;
- Compute the $c \times c$ covariance matrix Cov $= \bar{H}\bar{H}'$
- Compute the matrix E of eigenvectors which diagonalizes the covariance matrix Cov, i.e., $E^{-1}\mathrm{Cov}E = D$. The diagonal values of D are the eigenvalues associated to the eigenvectors that are column-wise arranged in E;
- Sort the columns of matrix E on the basis of the decreasing values of corresponding eigenvalues.

The principal components of T are the resulting eigenvectors.

Given an input vector $\mathbf{hc} \in \mathbb{R}^c$ it can be expressed as $\mathbf{hc} = \bar{\mathbf{hc}} + \sum_{j=1}^{c} \beta_j E_j$, where β_j is the projection of $\mathbf{hc}$ along the jth principal component, i.e., $\beta_j = (\mathbf{hc} - \bar{\mathbf{hc}}) \cdot E_j$. It is important to note that if the first $p < c$ principal components account for most of input data variability then an input vector $\mathbf{hc}$ can be expressed, with a small error only, as: $\mathbf{hc} \simeq \bar{\mathbf{hc}} + \sum_{j=1}^{p} \beta_j E_j$.

The outcome of the PCA analysis presented in Mason et al. (2001) shows that the first three eigenpostures (principal components are called eigenpostures in that context) account for more than 95% of hand-joints variability.

Eigenpostures appear to be particularly useful in hand grasping control. In fact, each hand configuration can be described by setting the coefficients of a linear combination of a small number of eigenpostures. More in general, one can reasonably assume that different classes of object-directed actions specified in terms of "type" and "modality" can be associated to different sets of eigenpostures in the space of hand-joints configurations. For example, two different sets of eigenpostures can be associated with "grasping with precision grip" and "grasping with whole hand prehension." These are two distinct classes of object-directed actions, coinciding with each other in the way of action type (grasping), and differing from each other in the way of their modality (whole hand vs. precision grip).

Formally, let r be the number of distinct classes of object-directed actions, and let $\mathcal{A}_k = \{E_j^k\}$ be the set of eigenpostures associated to the kth class, with $k = 1, 2, \ldots, r$, $j = 1, 2, \ldots, M_k$, and $E_j^k \in \mathbb{R}^c$, where c is the number of degrees of freedom. The eigenpostures associated to any class C span a low-dimensional sub-space of hand-joints configurations that a hand can assume during the execution of an object-directed action in C. Thus $\mathcal{A}_k$ is a basis for this subspace.

To sum up, a configuration $\mathbf{hc}$, that a hand can assume during the execution of an object-oriented action belonging to the kth class, can be described as:

$$\mathbf{hc} \simeq \bar{\mathbf{hc}} + \sum_{j=1}^{M_k} \beta_j \mathbf{E}_j^k \quad \text{with } \beta_j \in \mathbb{R} \tag{1}$$

From now on, a basis of eigenpostures is called *action subspace*. It is worth noting that if one knows the action subspace associated to a given class C of object-directed actions, then hand control during the execution of an action belonging to C can be achieved in a simplified fashion, insofar as it is sufficient to set a relatively small number of parameters. If one does not possess knowledge of the action subspace (equivalently, if this knowledge becomes temporarily unavailable), then the same hand control problem will require the setting of a more extended group of parameters, thus resulting into a less efficient solution.

Consistently with neurobiological models of F5 control functions, it is assumed that motor information encoded by mirror neuron activity is strictly related to action subspaces. In Sect. 4 below action subspaces are identified and associated to different classes of reach-to-grasp actions. Presently, let us turn to address the question of how motor information coded by mirror neuron activity can be deployed to drive the interpretation of perspectival sensory inputs. More specifically, this use of motor information in sensory processing is explored within a simplified action recognition scenario, wherein perspectival sensory inputs are restricted to hand-related visual inputs collected from some fixed viewpoint.

3.2 From motor to sensory processing

To begin with, consider a Perceptual System without motor information (P-Sys from now on), which computes scene descriptions from perspectival visual inputs only.

A major mathematical problem arising in this context concerns the ill-posed character of the required transformation from perspectival sensory data to intrinsic features of object-directed actions. To illustrate, suppose that the description one is looking for is given by a sequence of configurations that a hand takes on during an object-directed action. The same visually presented hand can be associated to various hand configurations, insofar as the hands of primates are highly complex structures including more than 20 degrees of freedom, and visually producing many different self-occlusions. Since visual sensory data collected from any given viewpoint are often compatible with various distal hand configurations, the mapping from perspectival visual inputs to hand configurations is a multi-valued function, thus assuming the form of an inverse ill-posed problem (Friston 2005; Fritsch 2007).

More precisely, let $X \subseteq \mathbb{R}^d$ be the d-dimensional space of visual hand descriptions, and let $Y \subseteq \mathbb{R}^c$ be the c-dimensional space of hand configurations. Then, one has to find a functional mapping f such that:

$$f : \mathbf{x} \in X \longrightarrow \wp(Y)$$

where $\wp(Y)$ is the power set of Y.

A two-dimensional example of a multi-valued function is illustrated in Fig. 4. This correspondence can be modelled by

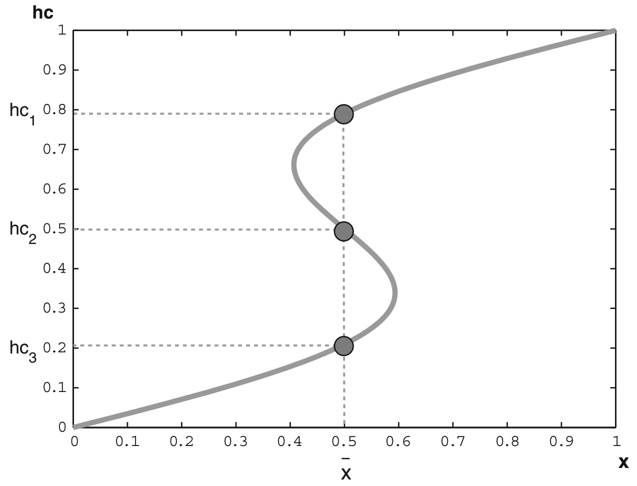


Fig. 4 Two-dimensional example of a multi-valued function. Points on the x-axis represent visual inputs, and points on the hc-axis represent hand-configurations. One may associate any given x point (visual input) with multiple hc points (hand-configurations). For example the $\bar{x}$ point is associated with points hc_1, hc_2, and hc_3

means of a probabilistic approach. More specifically, given $\bar{\mathbf{x}}$, the output computed by the mapping f can be approximated by the unconditional probability density function $p_{\bar{\mathbf{x}}}(\mathbf{hc})$ (see Fig. 4). Thus, in general, the problem of modelling the functional mapping f can be viewed in terms of estimating the conditional distribution $p(\mathbf{hc}|\mathbf{x})$.

According to Bishop (1995), one can cope with the problem of estimating $p(\mathbf{hc}|\mathbf{x})$, by adopting a Mixture Density Network (MDN) approach:

$$p(\mathbf{hc}|\mathbf{x}) = \sum_{i=1}^{M} \alpha_i(\mathbf{x})\phi_i(\mathbf{hc}|\mathbf{x}) \tag{2}$$

where the $\phi_i(\mathbf{hc}|\mathbf{x})$ are kernel functions that are usually identified with Gaussian functions of the form:

$$\phi_i(\mathbf{hc}|\mathbf{x}) = \frac{1}{(2\pi)^{c/2}\sigma_j^c(\mathbf{x})} \exp\left(-\frac{||\mathbf{hc} - \boldsymbol{\mu}_i(\mathbf{x})||^2}{2\sigma_i(\mathbf{x})}\right)$$

The parameters $\alpha_i(\mathbf{x})$ can be regarded as prior probabilities of $\mathbf{hc}$ generated from the ith component of the mixture $\phi_i(\mathbf{hc}|\mathbf{x})$.

The coefficients of the mixture, $\alpha_i(\mathbf{x})$, and the parameters of the kernel functions, $\phi_i(\mathbf{hc}|\mathbf{x})$, ($\boldsymbol{\mu}_i(\mathbf{x})$, and $\sigma_i(\mathbf{x})$ for a Gaussian kernel), depend on sensory input $\mathbf{x}$. A two-layer, feed-forward neural network can be used to model the relationship between visual inputs $\mathbf{x}$ and corresponding mixture parameters. In particular for an M components mixture model, the network will have $(c+2) \times M$ outputs.[1]

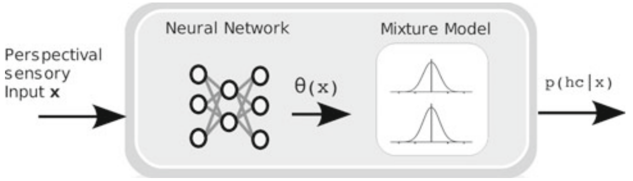


Fig. 5 Perceptual System (P-Sys). Mapping from perspectival input to direct internal input. A general framework for estimating conditional distribution $p(\mathbf{hc}(t_h)|\mathbf{x}(t_h))$ makes use of a density model and neural network combined structure

Accordingly, the problem of estimating the conditional probability distribution $p(\mathbf{hc}|\mathbf{x})$ can be approached by combining a density model and a neural network structure.

Thus, a system without motor information can be modeled as illustrated in Fig. 5. Given a previously unseen hand visual description $\mathbf{x}$, one can obtain an estimate of hand configuration $\mathbf{hc}$ by the following steps:

- the neural network component is fed with input $\mathbf{x}$;
- the network outputs are used to compute the mixture parameters α_i, $\boldsymbol{\mu}_i$, and σ_i, $i = 1, \ldots, M$;
- the distribution $p(\mathbf{hc}|\mathbf{x})$ is obtained as: $p(\mathbf{hc}|\mathbf{x}) = \sum_{i=1}^{M} \alpha_i \frac{1}{(2\pi)^{c/2}\sigma_j^c} \exp\left(-\frac{||\mathbf{hc}-\boldsymbol{\mu}_i||^2}{2\sigma_i}\right)$
- the hand configuration $\mathbf{hc}$ is obtained as the central value of the more probable branch of $p(\mathbf{hc}|\mathbf{x})$ (see Sect. 5 for more details).

Let us now introduce the proposed Motor-Perceptual System (MP-Sys from now on) which, unlike P-Sys, makes use of motor information.

Consistently with the discussion in the previous section, the existence of r distinct classes of object-directed actions is assumed, where each class C_k is associated with an action subspace $\mathcal{A}_k$. Furthermore, it is assumed that there is a selection mechanism which yields action subspaces $\mathcal{A}_1, \ldots, \mathcal{A}_r$ on the basis of (usually perceptual) information concerning the object towards which the action is directed. To each action subspace $\mathcal{A}_k$ one associates a probability $P(\mathcal{A}_k)$, and a prototypical/expected action in terms of a sequence $\bar{\boldsymbol{\beta}}^k(t_1), \bar{\boldsymbol{\beta}}^k(t_2), \ldots, \bar{\boldsymbol{\beta}}^k(t_m)$ which represents a "family" of actions, collectively expressed as a trajectory into the action subspace $\mathcal{A}_k$. Note that each action subspace corresponds to a specific class of object-directed actions, and that the values of the associated probabilities are initially set to prior probabilities P_i, i.e., $P(\mathcal{A}_i) = P_i$ with $i = 1, 2, \ldots, r$. These P_i's furnish an initial estimate of the probability of observing each class of object-directed actions. These probabilities are updated on the basis of the incoming visual input sequence $\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)$. More specifically, for each action subspace $\mathcal{A}_k$ and each time step t_h, one computes the probability that the prototypical/expected hand-configuration, coded as

[1] M outputs will determine the mixing coefficients $\alpha_1, \ldots, \alpha_M$, $M \times c$ outputs will determine the the components μ_{ij} of the kernel centers $\boldsymbol{\mu}_i$ with $i = 1, \ldots, M$ and $j = 1, \ldots, c$, finally M outputs will determine the kernel widths $\sigma_1, \ldots, \sigma_M$.

$\bar{\boldsymbol{\beta}}^k(t_h)$, corresponds to the actual hand-configuration on the basis of both incoming visual input $\mathbf{x}(t_h)$ and action subspace $\mathcal{A}_k$. These probabilities, denoted by π_h^k, are computed for the purpose of updating probabilities $P(\mathcal{A}_1), \ldots, P(\mathcal{A}_r)$. Accordingly, throughout an action recognition process, each $P(\mathcal{A}_k)$ value supplies a regularly updated estimate that the observed action is an object-directed action in the kth class.

In general, one can cope once again with the problem of estimating probabilities π_h^k by estimating at time t_h the probability distribution of $\mathbf{hc}(t_h)$, given the incoming sensory input $\mathbf{x}(t_h)$, i.e., $p(\mathbf{hc}(t_h)|\mathbf{x}(t_h))$. However, one can take advantage of knowledge concerning the action subspaces that are likely to be executed, by creating "action-based" distributions, $p_1(\mathbf{hc}(t_h)|\mathbf{x}(t_h))$, $p_2(\mathbf{hc}(t_h)|\mathbf{x}(t_h)), \ldots$, $p_r(\mathbf{hc}(t_h)|\mathbf{x}(t_h))$, one distribution for each action subspace. Moreover, in each action subspace, Eq. 1 enables one to express $\mathbf{hc}$ in terms of coefficients β_j. Thus, given sensory input $\mathbf{x}(t_h)$, one estimates the conditional probability distribution $p_k(\boldsymbol{\beta}(t_h)|\mathbf{x}(t_h))$ for each selected action subspace $\mathcal{A}_k$, and computes the π_h^k as $\pi_h^k = p_k(\bar{\boldsymbol{\beta}}^k(t_h)|\mathbf{x}(t_h))$.

In view of the fact that the information coded by mirror neurons activity is assumed to be strictly related to probabilities $P(\mathcal{A}_k)$, from now on $\text{mirror}_{\mathcal{A}_k}(t_h)$ is taken to denote the probability $P(\mathcal{A}_k)$ given the incoming visual input $\mathbf{x}$ at time t_h.

In this setting, given a previously unseen hand visual description $\mathbf{x}$, one can obtain an estimate of hand configuration in terms of $\boldsymbol{\beta}$, as the central value of the more probable branch of $p_k(\boldsymbol{\beta}(t_h)|\mathbf{x}(t_h))$ associated to the action subspace $\mathcal{A}_k$ with higher probability of $\text{mirror}_{\mathcal{A}_k}(t_h)$ (see Fig. 6).

During the observation of an object-directed action, the overall processing cycle of MP-Sys is summarized in terms of the processing steps of Algorithm 1. The same process is schematically illustrated by means of the functional diagram in Fig. 6.

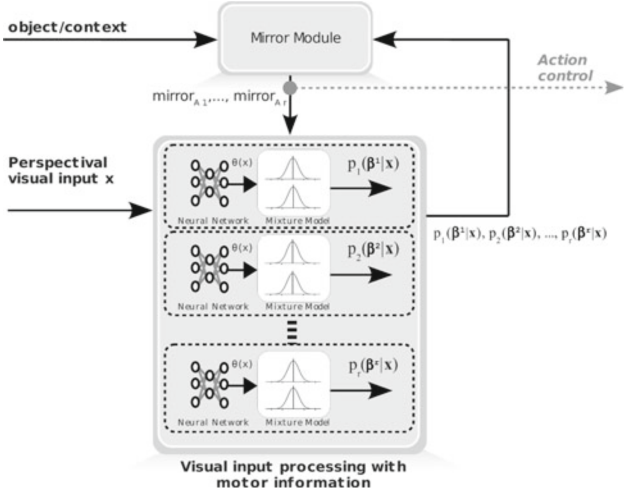

object/context
Mirror Module
Action control
$\text{mirror}_{A1}, \ldots, \text{mirror}_{Ar}$
Perspectival visual input x
θ(x)
Neural Network
Mixture Model
$p_1(\beta^1|x)$
$p_2(\beta^2|x)$
$p_r(\beta^r|x)$
$p_1(\beta^1|x), p_2(\beta^2|x), \ldots, p_r(\beta^r|x)$
Visual input processing with motor information

Fig. 6 Motor-Perceptual System (MP-Sys). Functional diagram of interactions between motor and perceptual processes in action observation (see text for details)

4 Action subspaces for different classes of grasp actions

A central presupposition of the computational model MP-Sys concerns the availability of different action subspaces (bases of eigenpostures) for different classes of object-directed actions. In this section, supporting evidence for the availability of these action subspaces is provided in connection with the grasping type of object-directed actions. More precisely, evidence is provided that different action subspaces for different modalities of grasping actions can be computed, by showing that action subspaces related to the same modality of reach-to-grasp actions "group together." The similarity measure proposed in Krzanowski (1979) is used to compare two different action subspaces (see next section for more details). Thus, we associate a basis of eigenpostures to each particular grasp modality, whereas in Santello et al. (2002) a basis of eigenpostures is isolated for the more generic class of grasp actions irrespectively of what is the grasp action modality.

4.1 Tests and results

The dataset of grasping action used in this experimental setting is more inclusive than the set of grasping actions that monkeys can perform (Gallese et al. 1996): it is a dataset

Algorithm 1 Action observation algorithm.

```
1. On the basis of (perceptual) information concerning the object towards which the action is directed, a set of action subspaces
   $\mathcal{A}_1, \ldots, \mathcal{A}_r$ is selected. Each selected action subspace $\mathcal{A}_k$ is associated to a probability value $\text{mirror}_{\mathcal{A}_k}(t_0) = P_k$
2. For $h \leftarrow 1$ to $m$ DO
   2.1 Let $\mathbf{x}(t_h)$ be the current visual input;
   2.2 On the basis of the selected action subspaces, generate prototypical hand-configurations coefficients $\bar{\boldsymbol{\beta}}^1(t_h), \ldots, \bar{\boldsymbol{\beta}}^r(t_h)$
   2.3 From the input $\mathbf{x}(t_h)$, compute probabilities $\pi_h^k = p_k(\bar{\boldsymbol{\beta}}^k(t_h)|\mathbf{x}(t_h))$ with $k = 1, \ldots, r$;
   2.4 On the basis of the computed probabilities $\pi_h^k$, update the probability values associated to each selected action subspace:
       $\text{mirror}_{\mathcal{A}_k}(t_h) \leftarrow \text{mirror}_{\mathcal{A}_k}(t_{h-1}) * \pi_h^k / (\sum_{i=1,\ldots,r} \text{mirror}_{\mathcal{A}_i}(t_{h-1}) * \pi_h^i)$
3. The final computed values $\text{mirror}_{\mathcal{A}_k}(t_m)$ identify the probability that the observed action is an action belonging to the kth class of
   actions.
```

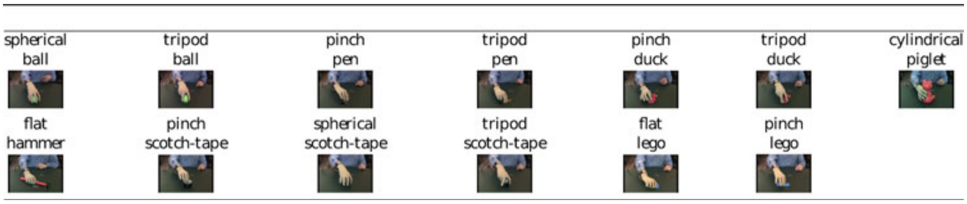


Fig. 7 The Grasp dataset used in this work was made available by *Lira Lab, University of Genova, Italy*. It is formed by grasping actions in five different modalities: cylindrical, spherical, tripod, flat, and pinch. The objects that these actions are directed to are: lego, scotch-tape, duck, piglet, hammer, pen, and ball. Some actions are directed toward more than one object

concerning human grasping actions (see Fig. 7) made available by *Lira Lab, University of Genova, Italy*. This dataset consists of grasping actions in five different modalities (*cylindrical, spherical, tripod, flat, and pinch*), some of which are directed towards different objects, and all of which are repeated 20 times by one subject. The pinch grasp is identical to a precision grip (PG) grasp; and every other grasp modality is a variant of the whole hand (WH) grasp. Figure 7 summarizes the set of grasping actions together with the objects toward which the actions are directed.

For each grasping action, the dataset contains the sequence of hand configurations obtained by means of the dataglove *CyberGlove* (*CyberGlove; Virtual Technologies, Palo Alto, CA, USA*) endowed with 22 sensors.

In order to compute appropriate bases for eigenpostures (action subspaces), a set of 13 different matrices $\{T^i\}_{i=1,\ldots 13}$ each containing all vectors **hc** (resulting from the combination of different grasping modalities and objects) was obtained and arranged row-wise. Each of these matrices contains data concerning actions of a specific grasping modality directed toward some specific object. Accordingly, each matrix T^i contains a number $c = 22$ of columns which is equal to the number of sensors, and a number of rows depending on the number of actions and their duration.

Finally, a PCA was performed on each matrix T^i. It turns out that four components suffice for capturing $\sim$ 90% of data variability for all modalities of grasping actions. Thus, an action subspace was obtained for each modality of grasping action and object; and the similarity measure between each pair of action subspaces was computed in accordance with (Krzanowski, 1979).

In particular, if T^i and T^j are two sets of data matrices and $\mathcal{A}_i$ and $\mathcal{A}_j$ are the corresponding action subspaces, both containing k eigenpostures (principal components), then the similarity measure is obtained by means of the following procedure:

- Let E^i and E^j be the matrices containing the column-wise arranged basis of eigenpostures of $\mathcal{A}_i$ and $\mathcal{A}_j$, respectively;
- Construct the $k \times k$ matrix G as: $G = L^T M M^T L$;
- Compute the k eigenvalues $\lambda_1, \ldots, \lambda_k$ of G;
- Obtain the similarity measure between $\mathcal{A}_i$ and $\mathcal{A}_j$ as $\text{sim}(\mathcal{A}_i, \mathcal{A}_j) = \sum_{i=1}^{k} \lambda_i$.

It can be shown that the sum of the eigenvalues of G equals the sum of squares of the cosines of the angles between each of the k eigenpostures defining action subspace $\mathcal{A}_i$ and each one of the eigenpostures defining $\mathcal{A}_j$. The similarity measure lies between 0 (orthogonal spaces) and k (coincident spaces).

In order to test whether there are different action subspaces, a divisive clustering (Hastie et al. 2003) between action subspaces was performed. The resulting clusters are shown in Figs. 8 and 9 obtained by setting threshold toll $=$ 0.75 and toll $=$ 0.7, respectively.

In the first case, only three clusters were obtained. It turns out that all pinch grasps are grouped together in *cluster* 2. In the same cluster one finds another kind of grasping action, i.e., a flat grasp directed toward a lego object. One should be careful to note, however, that the latter kind of action bears a strong similarity in hand shape with pinch actions in general, and with the *pinch lego* grasping action in particular. Moreover, the cylindrical grasping action, whose instances are directed toward the piglet toy object only, is placed into a different cluster. Finally, the remaining three classes of grasping actions (tripod, spherical, and flat) are grouped together in *cluster 1*.

One can achieve a better understanding of these clustering results by changing the $toll$ value. By setting toll $= 0.7$, one obtains four clusters (Fig. 9). As a result, *cluster 1* of the previous experiment is now split into two different clusters: *cluster 1* and *cluster 4*. If one compares the elements of each cluster, one cannot fail to note that different grasping actions in the same cluster share a common hand shape, depending on object shape features.

The above clustering results suggest the following considerations:

- different action subspaces exist for the two main modalities of grasping actions. Indeed, all pinch grasping actions (which are the same as precision grip grasping actions) group together. All other grasping actions, which are very similar to whole hand grasping actions, are basically grouped (except for cylindrical grasp) in a different cluster;

Fig. 8 Results of divisive clustering with toll = 0.75

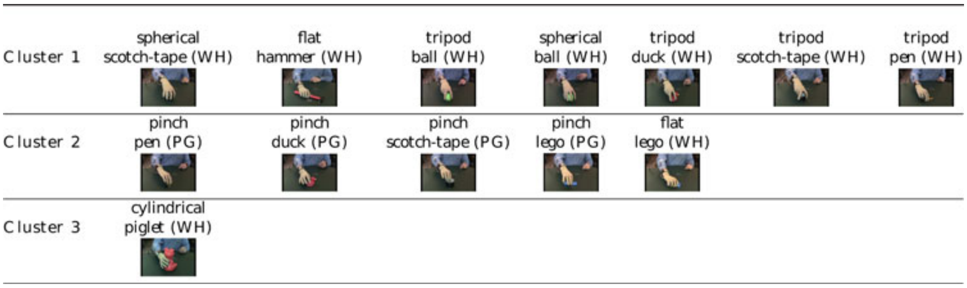
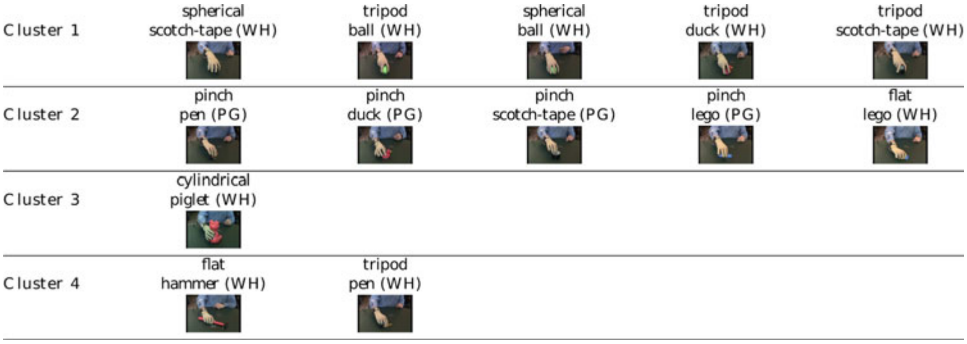


Fig. 9 Results of divisive clustering with toll = 0.7



- In everyday grasping actions there is a strict relation between grasping modality and the object toward which the action is directed (Napier 1956). By contrast, in the dataset some grasping actions are forced into an unusual modality with respect to the graspable object. For instance, the lego object is grasped in the dataset with a flat grasp too, even though one more usually grasps this object with a pinch grasp. Thus, distinguishing between different modalities of grasping actions appears to be a more difficult task in this dataset than in everyday life.

5 Motor information improves hand-configuration estimate

As explained in Sect. 3, motor information is treated here as a prior information supplied by the motor system to the visual system to interpret incoming visual inputs.

Now, we turn to show that one actually obtains more accurate estimates of hand-configurations from the observation of reach-to-grasp actions if one makes use of motor information taking the form of action subspaces. This demonstration is achieved by comparing MP-Sys and P-Sys described in Sect. 3.2 in their ability to estimate hand-configurations.

5.1 Experimental set-up

The goal of this test is to evaluate the benefit flowing from the use of motor information in estimates of hand-joints configurations during action observation.

To begin with, precision grip (PG) and whole hand (WH) grasping actions executed by a human being were recorded by means of the HumanGlove (*HumanGlove, Humanware S.r.l., Pontedera (Pisa), Italy*) endowed with 16 sensors. This dataglove feeds data into a 3D rendering software which reads sensor values and constantly updates a 3D human hand model. Thus, this experimental setting enables one to collect pairs *hand-joints configuration—hand image*. Twenty PG actions and twenty WH actions were recorded. Half of these were used to form the *training set* and the remaining ones were used as *test set*.

In order to extract vectors of visual features $\mathbf{x}$, each image of size 670×490 pixels is converted into a grayscale picture, subsampled at size 151×112 pixels and linearized into a single vector of size 1×16912. A PCA algorithm (Hastie et al. 2003) is applied over the dataset of collected hand images and the first five principal components only are computed. Each image is projected in the space of the first five principal components.

One may doubt that only five principal components suffice to obtain a good representation of data variability. There is, however, a 2-fold motivation for this choice. First, input images generated by a 3D rendering software are affected by little or negligible noise; therefore, the problem of estimating hand configuration from such images is drastically simplified with respect the problem of estimating hand configuration from real world images. Second, let us recall that we are concerned with the ill-posed problem discussed in Sect. 3.2. By taking the first five principal components only, one discards a considerable amount of input

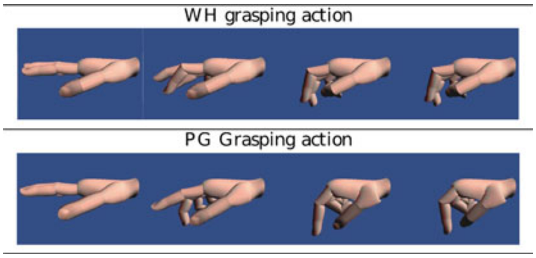


Fig. 10 A sample of extracted hand configurations with respect to the two different classes of actions

information, thus aptly making the problem of estimating hand configuration "more difficult" with respect to problems in which hand images are coded by many principal components.

The 3D simulator was prepared to synthesize hand configurations from just one fixed point of view, assumed by an observer who is asked to recognize correct hand configurations. Figure 10 shows sample pictures extracted from two different actions.

The appropriate action subspaces are computed for both classes PG and WH in the same way as described in Sect. 4.1. Three eigenpostures for each action subspace suffice to describe more than 90% of the total variability of hand-joints configurations **hc**. Thus, each hand-joints configuration is expressible in terms of the $\boldsymbol{\beta}$ coefficients.

Two MDN networks were used in connection with MP-Sys.

Each MDN was trained on data concerning one action class only (PG or else WH) to estimate $p_{\mathrm{PG}}(\boldsymbol{\beta}|x)$ and $p_{\mathrm{WH}}(\boldsymbol{\beta}|x)$, respectively. P-Sys is formed by one MDN network only, and is trained on the whole dataset (PG and WH) to estimate distribution $p(\mathbf{hc}|x)$.

The training process aims at determining the network parameters (network weights) by maximizing the following error function:

$$E = -\sum_{j=1}^{n}\left\{\sum_{i=1}^{M}\alpha_i(\mathbf{x}^j)\phi_i(\mathbf{hc}^j|\mathbf{x}^j)\right\} \tag{3}$$

This is achieved by means of a standard back-propagation algorithm, whose updating rules are described in Bishop (1995).

All MDNs were trained with different numbers H of hidden units and K of kernels for the neural network and mixture component, respectively. For each kernel and hidden unit configuration, the training was repeated T times. Only the best MDN configurations in terms of hidden units and number of kernels were selected. Table 1 summarizes the parameters used for tests.

Table 1 Experimental parameters

H	K	T	m
From 2 to 10 at step 2	From 5 to 20 at step 5	10	45

The two prototypical actions, as introduced in Sect. 3.2, are built out of sequences of hand-joints configurations coefficients as follows. Actions belonging to the training set of classes PG and WH, respectively, were used to obtain the two prototypical sequences $\bar{\boldsymbol{\beta}}^{\mathrm{WH}}(t_1), \ldots, \bar{\boldsymbol{\beta}}^{\mathrm{WH}}(t_m)$ and $\bar{\boldsymbol{\beta}}^{\mathrm{PG}}(t_1), \ldots, \bar{\boldsymbol{\beta}}^{\mathrm{PG}}(t_m)$.

The elements $\bar{\boldsymbol{\beta}}^{\mathrm{WH}}(t_j)$ and $\bar{\boldsymbol{\beta}}^{\mathrm{PG}}(t_j)$ at time t_j are obtained as:

$$\bar{\boldsymbol{\beta}}_k^{\mathrm{WG}}(t_j) = \frac{1}{M^{\mathrm{WH}}}\sum_{i=1}^{M^{\mathrm{WH}}}\boldsymbol{\beta}_i^{\mathrm{WH}}(t_j)$$

$$\bar{\boldsymbol{\beta}}_k^{\mathrm{PG}}(t_j) = \frac{1}{M^{\mathrm{PG}}}\sum_{i=1}^{M^{\mathrm{PG}}}\boldsymbol{\beta}_i^{\mathrm{PG}}(t_j)$$

i.e., the element of the sequence at time t_j is obtained as the mean of all coefficients of actions collected at time t_j. (Note that all recorded actions were aligned to a fixed length of m frames.)

The initial probability of each action subspace was fixed to: $\mathrm{mirror}_{\mathrm{PG}}(t_0) = 0.5$ and $\mathrm{mirror}_{\mathrm{WH}}(t_0) = 0.5$.

In order to compare MP-Sys with P-Sys the following procedure was adopted. For each action belonging to classes PG and WH, both systems receive as input a sequence of images $\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_m)$. And each $\mathbf{x}(t_j)$ belonging to the sequence is processed as follows:

- P-Sys computes $p(\mathbf{hc}(t_j)|\mathbf{x}(t_j))$ and the predicted hand-configuration $\widetilde{\mathbf{hc}}$ as the central value $\mu_k(\mathbf{x}(t_j))$ of the more probable branch of the distribution[2] $p(\mathbf{hc}(t_j)|\mathbf{x}(t_j))$;
- MP-Sys computes the two distributions $p_{\mathrm{PG}}(\boldsymbol{\beta}(t_j)|\mathbf{x}(t_j))$ and $p_{\mathrm{WH}}(\boldsymbol{\beta}(t_j)|\mathbf{x}(t_j))$, in addition to the corresponding $\mathrm{mirror}_{\mathrm{PG}}(t_h)$ and $\mathrm{mirror}_{\mathrm{WH}}(t_h)$ values. Only the distribution corresponding to the max between $\mathrm{mirror}_{\mathrm{PG}}(t_h)$ and $\mathrm{mirror}_{\mathrm{WH}}(t_h)$ is selected. Let us call this distribution simply $p(\boldsymbol{\beta}(t_h)|\mathbf{x}(t_h))$. The hand-configuration estimate for MP-Sys $\tilde{\boldsymbol{\beta}}(t_h)$ is again obtained as the central value $\mu_k(\mathbf{x})$ of the more probable branch of the $p(\boldsymbol{\beta}(t_h)|\mathbf{x}(t_h))$ distribution.

[2] In the case of mixture of the form expressed in Eq. 2, this value is easily obtained. In fact, since each component of the mixture model is normalized, assuming that the components of the distribution are well separated and have negligible overlap, the more probable branch is given by: $\arg\max \alpha_j(\mathbf{x})$.

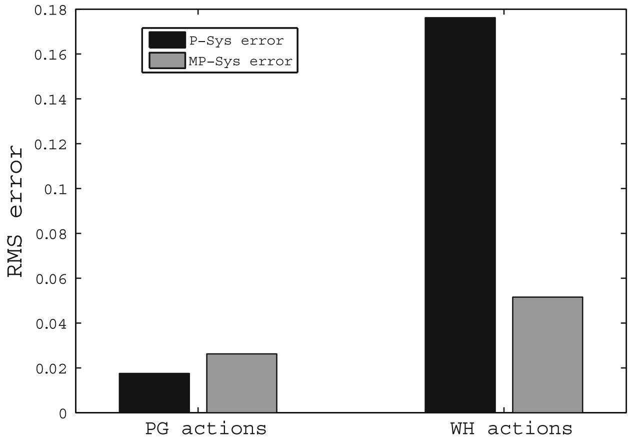


Fig. 11 Error over all actions in PG and WH classes is reported for both P-Sys and MP-Sys. Note that these systems achieve similar performances over the PG class. MP-Sys exhibits significantly better performances on the WH class

The error, for each hand configuration ($\tilde{\mathbf{hc}}$ and $\tilde{\boldsymbol{\beta}}$), is given by the Root Mean Square (RMS)[3] between actual visual input $\mathbf{x}$ and the predicted visual input $\mathbf{x}^{\tilde{\mathbf{hc}}}$ as well as between $\mathbf{x}$ and $\mathbf{x}^{\tilde{\boldsymbol{\beta}}}$ obtained by means of a forward model fed with $\tilde{\mathbf{hc}}$ and $\tilde{\boldsymbol{\beta}}$, respectively. A comment on the need for the forward model to compute the error is in order here. One may presume that such errors can be simply obtained as the RMS between computed hand configurations $\tilde{\mathbf{hc}}$ and $\tilde{\boldsymbol{\beta}}$ and expected configurations $\mathbf{hc}$ and $\boldsymbol{\beta}$ at time t_j. However, what we really want is a measure of how close are the $\tilde{\mathbf{hc}}$ and $\tilde{\boldsymbol{\beta}}$ to vectors representing real hand configurations that are compatible with the current visual input $\mathbf{x}$. Hence, by specifying a forward model which maps hand configurations to corresponding visual inputs one can compute the desired kind of measure.

5.2 Results

In Fig. 11 total errors of MP-Sys and P-Sys are reported. MP-Sys yields substantively less errors than P-Sys on WH actions. The error performances of these systems on PG actions are similar.

A better insight into the nature of the errors for both MP-Sys and P-Sys can be gleaned by looking at error trends during action observation. Consider first actions belonging to class WH. In Fig. 12, panels A, B, and C report data at each time step, averaged over all actions, while panels E, F, and G report data for each time and action.

Interestingly, from panels A and B (D and E), one notes that hand configuration prediction error for MP-Sys decreases when the value of mirror$_{\mathrm{WH}}$ correctly predicts what is the ongoing action (and increases otherwise). Moreover, by comparing panels B and C (E and F), one can see that MP-Sys errors are significantly lower than P-Sys errors when mirror$_{\mathrm{WH}}$ takes on higher values. A similar behavior was obtained for the PG action class, as shown in Fig. 13.

The above data suggest that the mirror module and visual modules interact with each other. However, one may still question the conclusion that it is motor information which improves or even “drives” visual processing. If this is indeed the case, one expects that correct predictions incoming from the mirror module will result into error decrease of MP-Sys.

This behavior of the mirror module can be forced from the outside by imposing that mirror$_{\mathrm{PG}} = 1$ (respectively, mirror$_{\mathrm{WH}} = 1$) at every time step and for any action of class PG (respectively, of class WH).

This expectation is corroborated by data in Fig. 14, showing an error decrease if the mirror module invariably provides correct feedback to the visual system. More specifically, panels A, B, and C report data concerning the WH action class; panels E, F, and G report data concerning the PG action class. Panels B and E show MP-Sys errors when mirror module feedback is as reported in panels A and D; panels C and F show MP-Sys errors when mirror module feedback is forced from the outside. Note that errors reported in panels C and F are significantly lower than errors reported in panels B and E, especially when mirror activity decreases (see panels A and D).

Finally, the accuracy achieved by MP-Sys for the two classes PG and WH is as follows: each observed action is assigned to class PG (respectively, WH) according to whether mirror$_{\mathrm{PG}}(t_m)$ is higher (respectively, lower) than mirror$_{\mathrm{WH}}(t_m)$. Every action in the test set for class WH was correctly classified; and 80% actions in the test set for class PG were correctly classified.

6 Discussion

A presupposition of the computational modelling approach presented above is the existence of a process enabling one to associate suitable prior probabilities P_i to each action subspace $\mathcal{A}_i$ on the basis of perceptual information concerning the object towards which the action is directed. In Tessitore et al. (2009) and Prevete et al. (2010), a biologically plausible computational system is introduced for affordance extraction in the context of grasping actions. This system appears to be readily modifiable for the purpose of computing these probabilities. One should be careful to note, however, that the results in Sect. 5 suggest the possibility of dispensing altogether with the process of extracting prior probabilities

[3] Root-mean-square error between target vectors $\mathbf{t}^n$ and model outputs $\mathbf{y}^n$ is computed as: $E^{\mathrm{RMS}} = \frac{\sum_{n=1}^{N} \|\mathbf{y}^n - \mathbf{t}^n\|^2}{\sum_{i=1}^{N} \|\mathbf{t}^n - \bar{\mathbf{t}}\|^2}$. Here, $\bar{\mathbf{t}}$ is defined to be the average target vector, that is: $\bar{\mathbf{t}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{t}^n$. The RMS has value one when the model predicts the test data “in the mean,” and value zero when the model’s prediction is perfect.

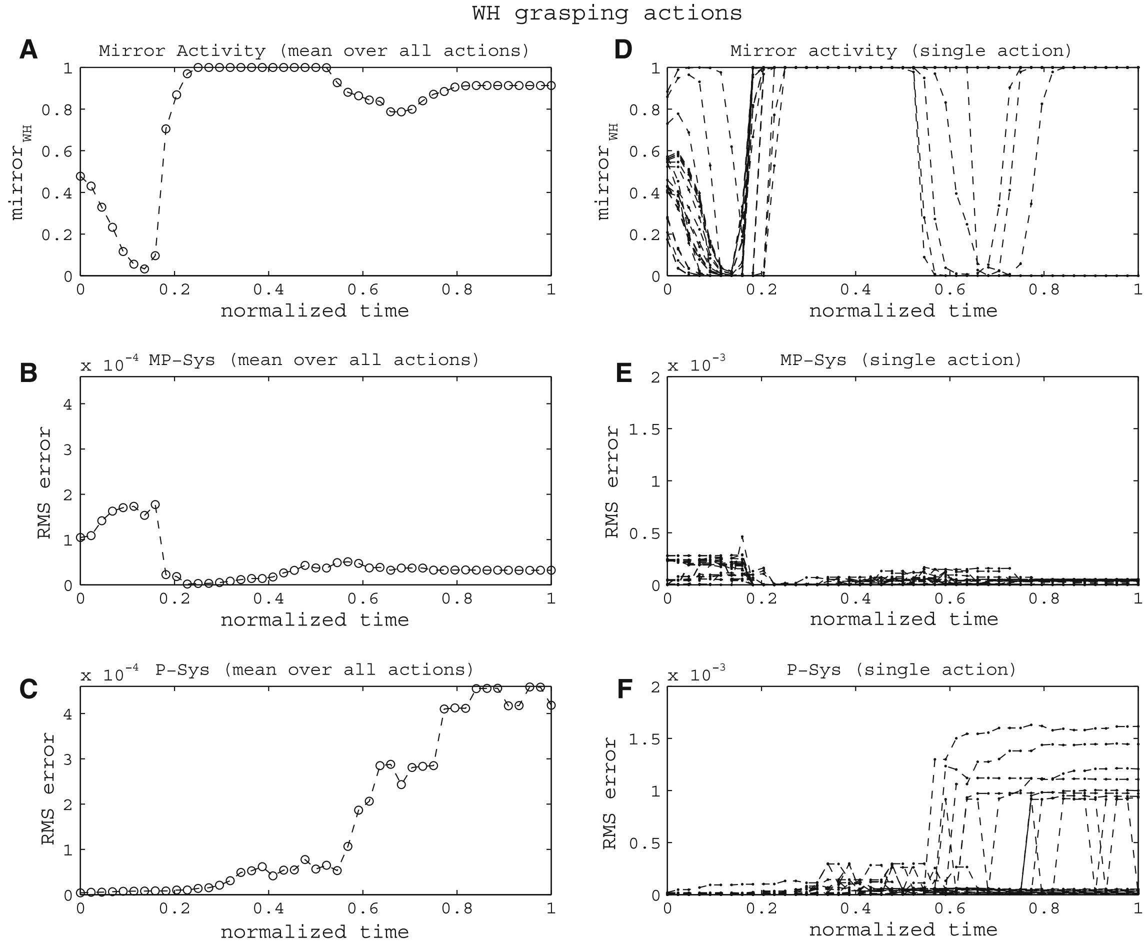


Fig. 12 Errors of P-Sys and MP-Sys together with mirror module outputs over the WH action class. Panels on the *left-hand side column* report data for each time step averaged over all actions; panels on the *right-hand side column* report data for each time and action. Note that MP-Sys errors follow a trend which is very similar to the output of the mirror module. Moreover, when the mirror module outputs the right prediction wrt the ongoing action, then MP-Sys errors are significantly lower than P-Sys errors

P_i. Indeed, if one sets to 0.5 the initial probability of each action subspace, as though no prior information were available, the system happens to respond correctly. Although this preliminary result suggests that extracting prior probabilities P_i might be unnecessary, additional theoretical and experimental investigations are needed to achieve a deeper understanding of this problem.

One may argue that the experimental set-up used to test systems P-Sys and MP-Sys is too simplified, because it includes two classes only of object-directed actions. However, it is reasonable to expect that the benefits of the proposed approach will be even more evident when larger numbers of object-directed action classes are taken into account. Indeed, the ill-posed character of the recognition problem discussed in Sect. 3.2 is likely to emerge more severely in these broader action contexts, in view of the increasing number of hand visual inputs which correspond to more than one hand configuration. In these broader action contexts, P-Sys has to deal with a more complicated distribution $p(\mathbf{hc}|\mathbf{x})$, presumably including higher numbers of distribution branches. In contrast with this, MP-Sys benefits from a priori information in terms of action subspaces and corresponding action-based visual processes. Each action-based visual process has to deal only with data concerning the associated object-directed action, and the corresponding distribution presumably has a limited number of branches.

Similar computational modelling approaches, allowing for iterative interactions between sensory processing and mirror activity, were explored in Haruno et al. (2001) and Ito and Tani (2004). This iterative interaction is modelled there by means of a series of computational modules implementing some sort of coupled forward-inverse models. The forward model predicts the next sensory stimuli on the basis of motor information; predicted sensory stimuli are compared with actual stimuli; and the outcome of this comparison enables the inverse models to update motor information. Then, by assuming that direct internal input is computed in a view-independent fashion, motor information

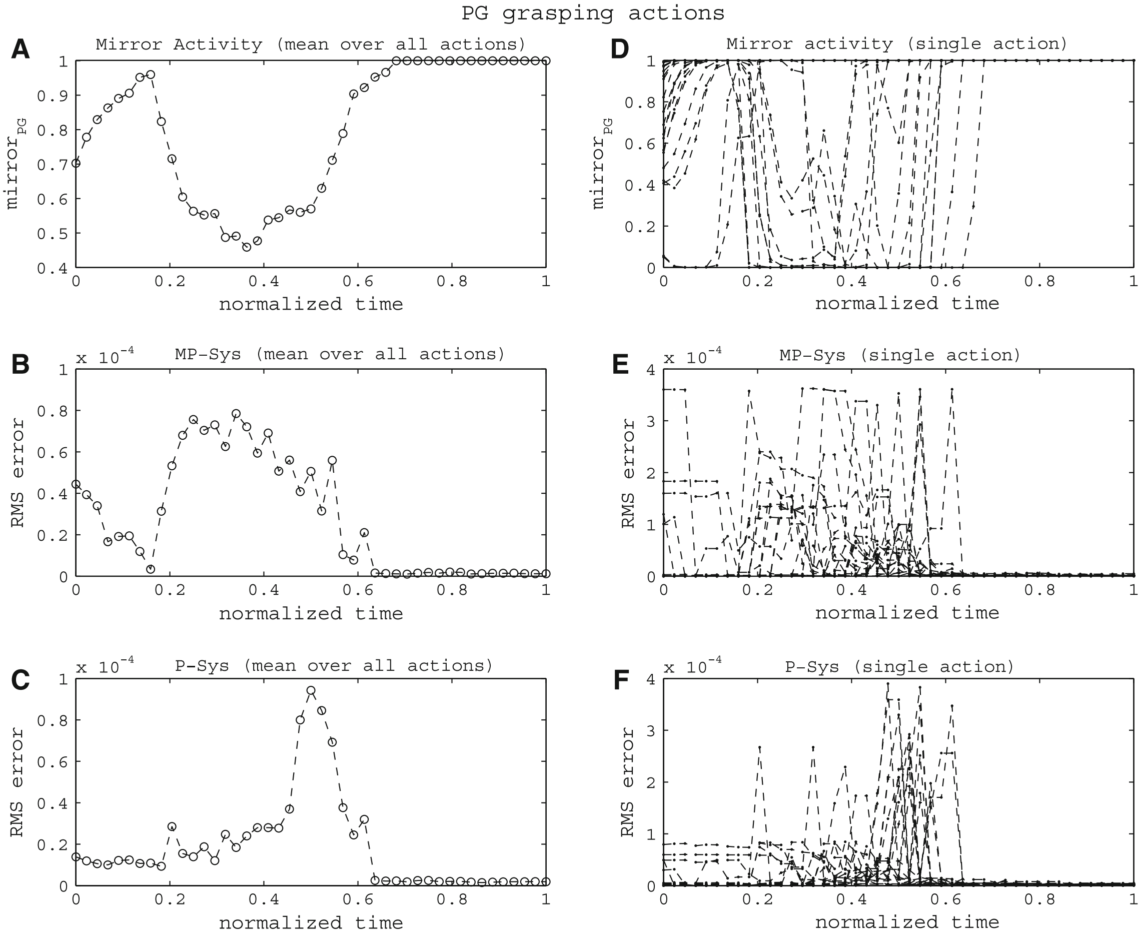


Fig. 13 Errors of P-Sys and MP-Sys together with mirror module outputs over PG action class. Panels on the *left-hand side column* report data for each time averaged over all actions; panels on the *right-hand side column* report data for each time and action. Note that MP-Sys errors follow a trend which is very similar to the output of the mirror module. Moreover, when the mirror module outputs the right prediction wrt the ongoing action, then MP-Sys errors are significantly lower than P-Sys errors

in action observation can be interpreted as mirror activity, insofar as this information coincides with motor information used to control a similar action. This overall picture of mirror mechanisms differs from the functional model presented here on various accounts. First, the present approach concentrates, in accordance with the direct matching hypothesis, on the use of motor information in sensory (visual) processes. In contrast with this, in current computational models the focus is shifted to obtaining the mirror property. Second, in our approach motor information "drives" visual processing (see the last experiment in the previous section), whereas in current computational models which include sensory-motor loop, motor information does not bear on visual processing. Finally, the sensory-motor loop modelled in Haruno et al. (2001) and Ito and Tani (2004) crucially involves a kinematic level of description for actions; in the present approach instead, motor information encodes action subspace probabilities which refer as a whole to action classes, without requiring a precise reference to action kinematic parameters. This approach to action representation is coherent with experimental data on F5 neurons in general, and F5 mirror neurons in particular, whose behavior is reported to be largely unselectively to the kinematic characteristics of executed/observed actions (Craighero et al. 2002).

In action execution mode, mirror neuron activity is part of a more complex neuronal activity including several cortical areas; furthermore, the behavior of mirror neurons is indistinguishable in execution mode from the behavior of other F5 motor neurons. The problem of how mirror neurons activity arises during action execution has not been examined here, but a functional interpretation of mirror neuron role in action execution has been nevertheless provided: motor information coded by mirror neurons is a simplifying factor in action control processes. The same motor information is used in MP-Sys to interpret perspectival sensory inputs. This use of motor information in the visual processing of perspectival inputs is the chief distinguishing feature of the present approach.

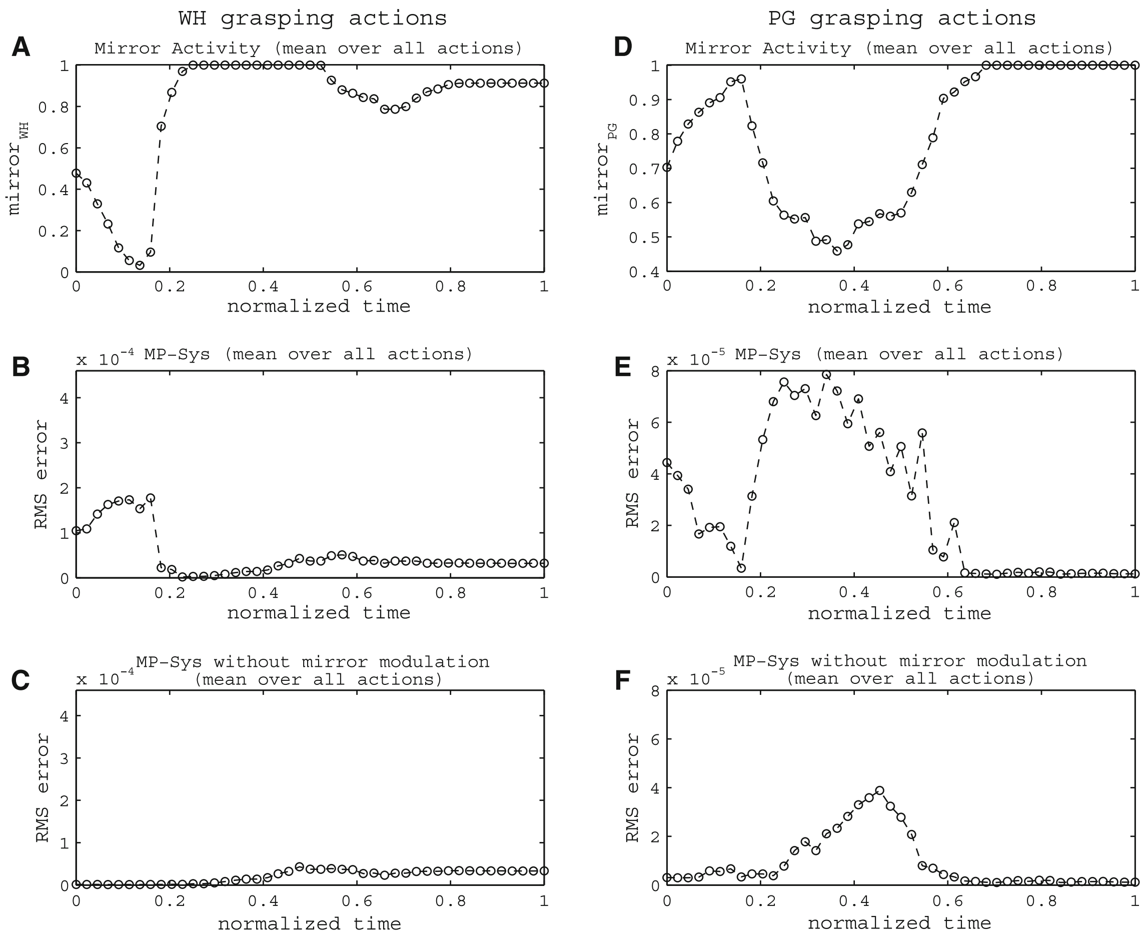


Fig. 14 Figure shows that motor information influences or "drives" visual processing. In fact, if motor information furnished by mirror activity is always correct in the way of ongoing action prediction, then error in hand-configuration estimate decreases

F5 mirror neurons are identified here with the functional units providing motor information codified in terms of action subspaces. And a candidate neuronal area for the execution of action-based visual processes driven by motor information (see Sect. 3.2) is temporal area STS. Indeed, the F5 mirror area is reciprocally connected to the inferior parietal lobule (IPL), and through IPL to STS (Matelli and Luppino 2001). Moreover, it is suggested in Keysers and Perrett (2004) and Oztop and Arbib (2002) that cortical area STS is involved in the process of extracting certain kinds of high-level visual features. In particular, "shape selective" cells were found in area STS which are highly selective for hand/object interactions exhibiting a mirror-like property (Keysers and Perrett 2004; Perrett et al. 1989). In fact, these neurons selectively respond to object-directed actions such as tearing, grasping, and manipulating. Unlike mirror neurons, however, shape-selective neurons fail to exhibit motor properties. Finally, let us note that STS neurons were hypothesized to provide a "pictorial description" of the ongoing action (Gallese and Goldman 1998), and their functional role was connected to high-level visual feature extraction processes whose outcomes are fed into mirror systems (Keysers and Perrett 2004). Taken together, these observations suggest the working hypothesis that action-based visual processes are chiefly located in cortical area STS, so that a cortical circuit between area STS and F5 mirror neurons plays a crucial supporting role for the visuo-motor loop functionally modeled and analyzed here.

Acknowledgments We wish to thank Corrado Sinigaglia for stimulating discussions. Support of the European Commission, VII FP (grant n. ICT-216293 DEXMART) and Compagnia di San Paolo, project *Action Representations and their Impairment* (2010–2012) under the Neuroscience Programme, is gratefully acknowledged.

References

Arbib MA (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. Behav Brain Sci 28(2):105–124

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Carr L, Iacoboni M, Dubeau MC, Mazziotta JC, Lenzi GL (2003) Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. PNAS 100(9):5497–5502

Cattaneo L, Rizzolatti G (2009) The mirror neuron system. Arch Neurol 66(5):557–560

Craighero L, Bello A, Fadiga L, Rizzolatti G (2002) Hand action preparation influences the responses to hand pictures. Neuropsychologia 40(5):492–502

Fleischer F, Casile A, Giese MA (2009) View-independent recognition of grasping actions with a cortex-inspired model. In: 9th IEEE-RAS international conference on humanoid robots, pp 514–519

Fogassi L, Gallese V, Buccino G, Craighero L, Fadiga L, Rizzolatti G (2001) Cortical mechanism for the visual guidance of hand grasping movements in the monkey. a reversible inactivation study. Brain 124:571–586

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B 360(1456):815–836

Fritsch C (2007) Predictive coding: an account of the mirror neuron system. Cognitive Processing 8:159–166

Gallese V, Goldman A (1998) Mirror neurons and the simulation theory of mind-reading. Trends Cognitive Sci 2(12):493–501

Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609

Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston

Haruno M, Wolpert DM, Kawato M (2001) Mosaic model for sensorimotor learning and control. Neural Comput 13(10):2201–2220

Hastie T, Tibshirani R, Friedman JH (2003) The elements of statistical learning: data mining, inference, and prediction, corrected edn. Springer

Iberall T, Fagg AH (1996) Neural network models for selecting hand shapes. In: Wing AM, Haggard P, Flanagan JR (eds) Hand and brain: neurophysiology and psychology of hand movements. Academic Press, San Diego, CA, pp 243–264

Iberall T, Bingham G, Arbib MA (1986) Opposition space as a structuring concept for the analysis of skilled hand movements, vol 15. In: Heuer H, Fromm C (eds) Generation and modulation of action patterns. Springer, Berlin, pp 158–173

Ito M, Tani J (2004) Generalization in learning multiple temporal patterns using RNNPB. In: ICONIP: International conference on neural information processing, pp 592–598

Keysers C, Perrett DI (2004) Demystifying social cognition: a hebbian perspective. Trends Cogn Sci 8(11):501–507

Krzanowski WJ (1979) Between-groups comparison of principal components. J Am Stat Assoc 74(367):703–707

Mason CR, Gomez JE, Ebner TJ (2001) Hand synergies during reach-to-grasp. J Neurophysiol 86(6):2896–2910

Matelli M, Luppino G (2001) Parietofrontal circuits for action and space perception in the macaque monkey. NeuroImage 14:S27–S32

Miall RC (2003) Connecting mirror neurons and forward models. Neuroreport 14(17):2135–2137

Napier JR (1956) The prehensile movements of the human hand. J Bone Joint Surg 38B:902–913

Oztop E, Arbib MA (2002) Schema design and implementation of the grasp-related mirror neuron system. Biol Cybernet 87:116–140

Oztop E, Wolpert DM, Kawato M (2005) Mental state inference using visual control parameters. Cogn Brain Res 22(2):129–151

Oztop E, Kawato M, Arbib MA (2006) Mirror neurons and imitation: a computationally guided review. Neural Netw 19:254–271

Perrett DI, Harries MH, Bevan R, Thomas S, Benson PJ, Mistlin AJ, Chitty AJ, Hietanen JK, Ortega JE (1989) Frameworks of analysis for the neural representation of animate objects and actions. J exp Biol 146:87–113

Prevete R, Tessitore G, Santoro M, Catanzariti E (2008) A connectionist architecture for view-independent grip-aperture computation. Brain Res 1225:133–145

Prevete R, Tessitore G, Catanzariti E, Tamburrini G (2010) Perceiving affordances: a computational investigation of grasping affordances. Cogn Syst Res. doi:10.1016/j.cogsys.2010.07.005

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Ann Rev Neurosci 27:169–192

Rizzolatti G, Sinigaglia C (2010) The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat Rev Neurosci 11:264–274

Rizzolatti G, Fadiga L, Gallese V, Fogassi L (1996) Premotor cortex and the recognition of motor actions. Cogn Brain Res 3(2):131–141

Rizzolatti G, Fogassi L, Gallese V (2001) Neurophysiological mechanisms underlying the understanding and imitation of action. Nat Rev Neurosci 2(9):661–670

Santello M, Flanders M, Soechting JF (2002) Patterns of hand motion during grasping and the influence of sensory guidance. J Neurosci 22(4):1426–1435

Tessitore G, Borriello M, Prevete R, Tamburrini G (2009) How direct is perception of affordances? A computational investigation of grasping affordances. In: Howes RC, Peebles D (eds) 9th international conference on cognitive modeling—ICCM2009, Manchester, UK

Umilta' M, Kohler E, Gallese V, Fogassi L, Keysers C, Rizzolatti G (2001) I know what you are doing: a neurophysiological study. Neuron 31(19):155–165