Human Movement Science 65 (2019) 121–141

Contents lists available at ScienceDirect

Human Movement Science

journal homepage: www.elsevier.com/locate/humov

Full Length Article

Fronto-parietal mirror neuron system modeling: Visuospatial transformations support imitation learning independently of imitator perspective

Hyuk Oh[a,b], Allen R. Braun[a,c], James A. Reggia[a,d,e], Rodolphe J. Gentili[a,b,f,*]

[a] Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD 20742, USA
[b] Department of Kinesiology, University of Maryland, College Park, MD 20742, USA
[c] Behavioral Biology Branch, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA
[d] Department of Computer Science, University of Maryland, College Park, MD 20742, USA
[e] University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA
[f] Maryland Robotics Center, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Keywords:
Mirror neuron system
Cognitive-motor learning
Visuospatial processing
Action imitation
Internal models
View-based processes

ABSTRACT

Although the human mirror neuron system (MNS) is critical for action observation and imitation, most MNS investigations overlook the visuospatial transformation processes that allow individuals to interpret and imitate actions observed from differing perspectives. This problem is not trivial since accurately reaching for and grasping an object requires a visuospatial transformation mechanism capable of precisely remapping fine motor skills where the observer's and imitator's arms and hands may have quite different orientations and sizes. Accordingly, here we describe a novel neural model to investigate the dynamics between the fronto-parietal MNS and visuospatial processes during observation and imitation of a reaching and grasping action. Our model encompasses i) the inferior frontal gyrus (IFG) and inferior parietal lobule (IPL), regions that are postulated to produce neural drive and sensory predictions, respectively; ii) the middle temporal (MT) and middle superior temporal (MST) regions that are postulated to process visual motion of a particular action; and iii) the superior parietal lobule (SPL) and intra-parietal sulcus (IPS) that are hypothesized to encode the visuospatial transformations enabling action observation/imitation based on different visuospatial viewpoints. The results reveal that when a demonstrator executes an action, an imitator can reproduce it with similar kinematics, independently of differences in anthropometry, distance, and viewpoint. As with prior empirical findings, similar model synaptic activity was observed during both action observation and execution along with the existence of both view-independent and view-dependent neural populations in the frontal MNS. Importantly, this work generates testable behavioral and neurophysiological predictions. Namely, the model predicts that i) during observation/imitation the response time increases linearly as the rotation angle of the observed action increases but remain similar when performing both clockwise and counterclockwise rotation and ii) IPL embeds

Abbreviations: AU, Arbitrary Unit; CB, Cerebellum; FWD, Forward model; GCV, Generalized Cross-Validation; IFG, Inferior Frontal Gyrus; INV, Inverse model; IPL, Inferior Parietal Lobule; IPS, Intra-Parietal Sulcus; MNS, Mirror Neuron System; MST, Middle Superior Temporal; MT, Middle Temporal; OLS, Orthogonal Least Square; RBF, Radial Basis Function; RMSE, Root-Mean-Square Error; rPFC, rostral Pre-Frontal Cortex; SPL, Superior Parietal Lobule; STS, Superior Temporal Sulcus; VD, View-Dependent; VI, View-Independent; VITE, Vector-Integration-to-Endpoint; VST, Visuospatial Transformation

* Corresponding author at: Cognitive Motor Neuroscience Laboratory, Department of Kinesiology, University of Maryland, School of Public Health (Bldg #255), Room #2138, College Park, MD 20742, USA.
E-mail address: rodolphe@umd.edu (R.J. Gentili).

https://doi.org/10.1016/j.humov.2018.05.013
Received 13 December 2017; Received in revised form 15 May 2018; Accepted 25 May 2018
Available online 12 September 2018
0167-9457/

essentially view-independent neurons while SPL/IPS includes both view-independent and view-dependent neurons. Overall, this work suggests that MT/MST visuomotion processes combined with the SPL/IPS allow the MNS to observe and imitate actions independently of demonstrator-imitator spatial relationships.

1. Introduction

A substantial number of human neuroimaging studies support the existence of a large brain network associated with sensorimotor integration called the *mirror neuron system* (MNS) (Aziz-Zadeh & Ivry, 2009; Bisio et al., 2015; Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Case, Pineda, & Ramachandran, 2015; Dinstein, Gardner, Jazayeri, & Heeger, 2008; Gueugneau, Bove, Ballay, & Papaxanthis, 2016; Gueugneau, McCabe, Villalta, Grafton, & Della-Maggiore, 2015; Iacoboni, 2005; Iacoboni et al. 1999; Iacoboni et al., 2001; Oztop, Kawato, & Arbib, 2013). The human MNS becomes active for both observation and imitation of actions, and may correspond to the mirror neurons identified in monkeys, although the human MNS responds to both the action (e.g., a reaching and grasping action) and the movements *per se* (e.g., arm reaching, finger pre-shaping, and finger closing movements) whereas the mirror neurons in primates are engaged only when observing the action as a whole (Buccino et al., 2001; Oztop et al., 2013; Rizzolatti, 2005; Rizzolatti, Fogassi, & Gallese, 2002; Rizzolatti & Craighero, 2004). In particular, experiments have revealed that the frontal (inferior frontal gyrus or IFG) and parietal (inferior parietal lobule or IPL) components of the human MNS are active during reaching to grasp action imitation (Grèzes & Decety, 2001; Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fogassi, & Gallese, 2001). MNS-like systems such as the superior temporal sulcus (STS) and temporo-parieto-frontal circuits are also involved during reaching to grasp action observation and imitation (Carr et al., 2003; Iacoboni et al., 2001).

In addition, a naïve spectator can easily see that when one person (a demonstrator) shows an action to another individual (an imitator), the latter can observe and reproduce these actions independently of spatial constraints such as differences in viewpoint (or perspective), distance, and anthropometry involving the two protagonists. However, while both experimental and computational studies have examined the neural mechanisms for action observation and imitation, only very limited attention has been given to the visuospatial processing that could account for such spatial differences between the demonstrator and the imitator. Specifically, in a typical imitation task, the respective frame of reference of the demonstrator (allocentric) and imitator (egocentric) is different (e.g., in orientation and/or position). Thus, to successfully imitate an observed reaching to grasp action, the imitator needs to transform (or remap) the observed arm and hand movement from the demonstrator's allocentric frame of reference into their egocentric frame of reference. This frame of reference transformation can be quite complex due to discrepancies in anthropometry and functional ranges of motion of the two protagonists. Indeed, this problem is not trivial, since the performance of a movement to reach and grasp an object accurately will require a frame of reference transformation able to precisely remap such relatively fine motor skills where the observed arm and hand have different orientations and size for observer/imitator. Only a few MNS studies have examined this perspective problem during action observation and imitation. Early non-human primate studies revealed that mirror neurons discharged similarly when the observed hand had various orientations. For instance, it was shown that observing grasping actions from various perspectives resulted in the same activation of the frontal MNS (e.g., Caggiano et al., 2011; Caggiano, Giese, Thier, & Casile, 2015; Gallese et al., 1996). In particular, examination of mirror neuron responses in primates while manipulating the viewpoint from which the grasping actions were observed revealed the existence of two mirror neuron populations in the frontal gyrus (F5), a view-dependent and a view-independent population (Caggiano et al., 2011). A few empirical human studies also examined the relationships between the viewpoint and action observation. These investigations found that actions observed with an egocentric viewpoint elicited larger brain responses in the contralateral IFG region similar to self-intended action. In contrast, actions observed from an allocentric viewpoint led to greater ipsilateral region activation (Shmuelof & Zohary, 2008). Other work has suggested possible modulatory mechanisms of the perspective (first or third person) on cross-modal action coding in the frontal MNS while the activity of the parietal MNS was found to be independent of the perspective (Oosterhof, Tipper, & Downing, 2012).

Altogether these studies suggest that, both in primates and humans, mirror neurons i) have a response that is modulated by the perspective from which the action is observed, and ii) interact with other neural structures that mediate action understanding by matching visual and motor representations of the observed actions. However, our current understanding of how such viewpoint processing interacts with the MNS is quite limited and requires further examination. The understanding of such visuospatial processing, as with any other mechanisms underpinning the MNS, will further inform the functions hypothesized to be attributable to the MNS (Oztop et al., 2013).

In addition to empirical efforts, several conceptual and computational models have been proposed to help understand the underlying functional roles of the MNS (Bonaiuto, Rosta, & Arbib, 2007; Miall, 2003; Oh, Gentili, Reggia, & Contreras-Vidal, 2012; Oztop, Wolpert, & Kawato, 2005; Oztop et al., 2013). Specifically, to model the MNS functionalities from a motor control and learning standpoint, a number of these models employed an internal model framework (e.g., Crevecoeur, Giard, Thonnard, & Lefevre, 2011; Crevecoeur & Scott, 2014; Jordan & Wolpert, 1999; Molina-Vilaplana & Coronado, 2006; Molina-Vilaplana, Feliu-Batlle, & López-Coronado, 2007; Wolpert & Miall, 1996). In particular, it has been suggested that the STS-IPL-IFG pathway implements an inverse model to encode the motor command from an observed action that is used for subsequent imitation (Miall, 2003). Also, the IFG-IPL-STS pathway has been hypothesized to implement a forward model that predicts sensorimotor states by using an efference copy of the motor command that will be subsequently used for action imitation (Miall, 2003). This model includes a cerebellar structure embedding other forward and inverse models in parallel with the temporo-parieto-frontal circuit (Miall, 2003).

These previous internal model framework efforts did not generally include a neural component able to mimic the visuospatial computational mechanisms that would transform the frames of reference between the demonstrator and the imitator (Arie, Arakaki, Sugano, & Tani, 2012; Bonaiuto & Arbib, 2010; Bonaiuto et al., 2007; Demiris & Hayes, 2002; Demiris & Johnson, 2003; Demiris & Khadhouri, 2006; Oztop & Arbib, 2002; Oztop, Bradley, & Arbib, 2004; Oztop, Kawato, & Arbib, 2006; Oztop et al., 2005, 2013). Only a very few past computational studies have attempted, to some extent, to account for differences of viewpoint between the demonstrator and imitator. For instance, Lopes and Santos-Victor (2005) included perspective transformations in their model, but their solution was purely mathematical with no biological relevance or visuospatial learning. Also, although it was not the primary focus, a more recent MNS computational study proposed a neural model that could process very limited angular changes between the imitator and the demonstrator (Arie et al., 2012). The authors suggested that to account for larger angular variations, a model of visuospatial processes such as mental rotation should be considered.

Still in a context of action observation/imitation, additional prior modeling efforts have proposed a neurocomputational mechanism for which the vector population, particularly in the STS, would play a critical role in implementing a transformation of the frame of reference between the demonstrator and imitator (Sauser & Billard, 2005a, 2005b, 2007). More recently, another neural model has examined the view-based mechanisms and was able to reproduce various neurophysiological findings observed in primates (Fleischer, Caggiano, Thier, & Giese, 2013). Although interesting, this previous modeling work did not: i) model the relationships between the mechanisms underlying the transformation of the frame of reference and the frontal and parietal MNS, ignoring thus the interactions between visual processing of reaching to grasp actions and motor representations; ii) consider the simultaneous learning of such transformation mechanisms with the fronto-parietal MNS in an ecological context where observation and imitation are intermingled, iii) examine, as a result of learning, the emergence of view-based representation in the frontal-parietal MNS and the neural substrate possibly embedding such transformation mechanisms and iv) consider scaling of the physical features (e.g., arm and hand) between demonstrator and imitator.

Both the experimental and computational works summarized above suggest that there is a critical need to further understand how visuospatial processes interact with the MNS. No previous neurocomputational model has examined the general functional mechanisms that underlie observing and imitating actions independently of differences in the imitator's viewpoint, the imitator's distance, and the imitator's body size relative to the demonstrator (Arie et al., 2012; Caggiano et al., 2011; Lopes & Santos-Victor, 2005; Sauser & Billard, 2005a, 2005b, 2007; Fleischer et al., 2013). The development of a neurocomputational model that combines visuospatial transformations with the MNS will inform the underlying working principles of the MNS and in particular will complement and extend prior work by examining how the view-based and sensorimotor representations interact during action observation and imitation. In addition, this work could also contribute to developing more efficient imitation learning mechanisms for humanoid robots (Floreano, Ijspeert, & Schaal, 2014; Gentili, Oh et al., 2015; Katz, Huang, Hauge, Gentili, & Reggia, 2017; Schaal & Schweighofer, 2005).

Here we present and study a neurocomputational model of the parietal visuospatial transformation processes that interact with the fronto-parietal MNS. Our model learns to perform observation and imitation of reaching and grasping actions regardless of visuospatial differences (viewpoint, distance, and anthropometry) between a demonstrator and an imitator. Our model aims to address the following questions. What are the computational processes underlying the visuospatial transformation mechanisms, and what is the role of such mechanisms with respect to the fronto-parietal MNS? What possible neural structure(s) could implement these visuospatial transformation mechanisms, and how would these structures interact with those from the fronto-parietal MNS? What behavioral and neural dynamics features will emerge from the interactions between the visuospatial transformation and the fronto-parietal MNS mechanisms during action observation and imitation? In particular, we hypothesize that by incorporating visuospatial transformation mechanisms that interact with the fronto-parietal MNS network, our neural model will not only produce similar arm and hand kinematics for a reaching and grasping action previously observed from various perspectives, but that it will also generate comparable neural activity in IFG and IPL during both action observation and imitation, and that view-dependent and view-independent neural populations reminiscent of those observed biologically will emerge spontaneously.

2. Methods

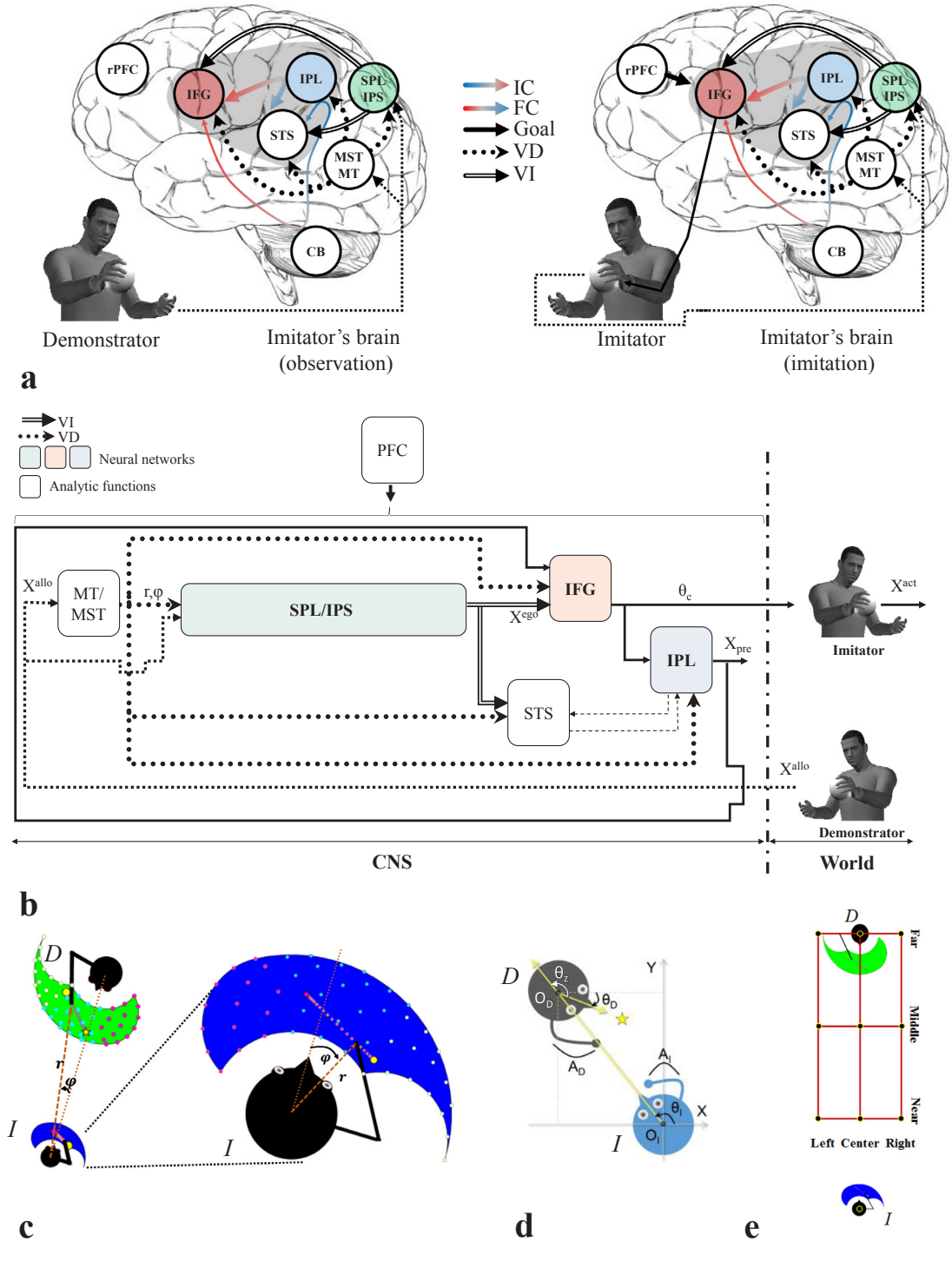
Here we describe the neural mechanisms and computations of our fronto-parietal model, their implementation and learning mechanisms, as well as the methods used to assess the performance of the model. In the following, "observation" refers to the observation phase of an action where no overt movement are produced, and "imitation" refers to the execution phase where the previously observed action is imitated through actual movements.

2.1. Neural mechanisms of the model

Our neural model was developed within the framework that links the concept of internal models to the MNS (Iacoboni et al., 1999; Miall, 2003). In particular, Miall has suggested that the STS-IPL-IFG and the IFG-IPL-STS pathways would implement an inverse model that computes the neural drive for action imitation and a forward model that predicts sensorimotor states, respectively (Miall, 2003). Here, we expand Miall's conceptual model by including the three following neural substrates. First, the middle temporal (MT) and middle superior temporal (MST) regions were incorporated to selectively process visual motion direction and velocity from a perceived action (Adelson & Movshon, 1982; Born & Bradley, 2005; Duijnhouwer, Noest, Lankheet, Van Den Berg, & Van Wezel, 2013; Tootell et al., 1995). The MT/MST are postulated to provide the MNS with a view-dependent representation of an observed action. Second, the brain regions superior parietal lobule (SPL) and and intraparietal sulcus (IPS) were included to

implement visuospatial transformation mechanisms that are postulated to provide the MNS with view-independent representation of an observed action (Andersen, 1987; Buneo & Andersen, 2006; Grefkes & Fink, 2005; Rizzolatti, Luppino, & Matelli, 1998). Thus, the SPL/IPS would allow the MNS to engage in robust action observation and imitation independently of visuospatial constraints. Third, in our model the rostral prefrontal cortex (rPFC) triggers the intention to imitate (Burgess, Dumontheil, & Gilbert, 2007; Dove, Pollmann, Schubert, Wiggins, & von Cramon, 2000; Meyer et al., 1997; Rogers et al., 1998).

From a computational standpoint, compared to other MNS models, our model differs in that it is able to capture the dynamics



(caption on next page)

Fig. 1. Conceptual model of the imitator's neural processes and the view-based representations. (a) Brain regions and pathways engaged during action observation (left panel) and imitation (right panel) (based on Miall, 2003) where the MT/MST and SPL/IPS were added. Only the connections relevant to our neural model are shown. The gray area represents the main MNS structures. The view-dependent information (thin dotted lines) provided by the visual cortex (not shown) are sent to the MT/MST and SPL/IPS. The MNS performs the inverse (IC) and forward (FC) computation via the STS-IPL-IFG and the reverse pathway, respectively. (b) Imitator's model brain areas and neural pathways involved during action observation and imitation. SPL/IPS, IPL, and IFG are implemented with neural networks and functionally mimic the hypothesized visuospatial transformations, forward and inverse computations, respectively. VD: view-dependent signals modulated by imitator's perspective (dotted lines); VI: view-independent signals not modulated by imitator's perspective (double solid lines). (c) Geometric modeling to encode the view-dependent representation in MT/MST. Left panel: The imitator first observes an action performed by a demonstrator. The view-dependent representation of the action (r, φ) of the demonstrator's effector (allocentric) from the imitator's viewpoint. Right panel: When the imitator observes the imitator's own action from an initial (yellow circle) to a final (magenta star) position, the view-dependent information (r, φ) is now encoded in its own egocentric frame of reference. This encoding process allows remapping the left, center, right (white, cyan, magenta circles) sides of the demonstrator into the left, center, right (magenta, cyan, white circles) sides of the imitator from the imitator's viewpoint. (d) Geometric modeling of the different origin of the frame of reference (O_I and O_D) and viewpoint (i.e., lines of sight) for each protagonist. The imitator observes a reaching action towards an object (yellow star) performed by the demonstrator. X and Y represent the coordinate axes of the absolute coordinate system. θ_I, θ_D: angles towards the imitator or the demonstrator's line of sight from the X-axis; θ_z: the rotation angle from the demonstrator to the imitator's viewpoint. (e) Areas of the environment where the demonstrator is located when showing actions to the imitator. Three orientations (imitator's left, front and right sides) and three demonstrator-imitator distances (near, middle and far) are used. Here the demonstrator is showing an action in front of and far away from the imitator. D: demonstrator; I: Imitator; (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

between view-based representations and frontal-parietal MNS. Specifically, taking inspiration from previous modeling work, the overall novelty of our model is that it examines frontal (adaptive inverse model) and parietal (adaptive forward scheme) MNS dynamics as well as their relationships with the SPL/IPS (adaptive visuospatial transformation that enable view-independent representation) and the MT/MST regions (view-dependent representation) during learning by observation and imitation (Oh, Gentili, Reggia, & Contreras-Vidal, 2011; Oh et al., 2012) (Fig. 1(a),(b)).

2.2. Neurophysiological components

Here we summarize the main neurophysiological systems that provide inspiration for the neurobiological components of our model.

The rPFC plays a critical role in many cognitive-motor functions, but here it is modeled as a simple top-down behavioral control that allows the mirror system to achieve goal-directed actions via switching between action observation and action imitation (Burgess et al., 2007; Miller & Cohen, 2001).

The STS region is reciprocally connected to IPL, and thus provides sensory (particularly visual) input to the MNS (Barraclough, Keith, Xiao, Oram, & Perrett, 2009; Kilner, Friston, & Frith, 2007; Miall, 2003). Also, since the STS responds to biological motion of body parts such as arms and hands (Kilner et al., 2007; Miall, 2003), it has been suggested that it has a role in action imitation via the temporo-parieto-frontal network (Rizzolatti et al., 2001). Thus, in our model the STS encodes hand/arm movements associated with specific actions such as reaching and grasping, while being able to recognize if the observed actions are already in the current motor repertoire of the imitator or not.

The cerebellum (CB) is also critical for sensorimotor error processing such as sensory prediction errors during motor learning (Han, Arbib, & Schweighofer, 2008; Ito, 2002; Popa, Hewitt, & Ebner, 2012; Schweighofer, Lang, & Kawato, 2013; Tokuda, Han, Aihara, Kawato, & Schweighofer, 2010). Thus, here the cerebellum was modeled as a simple comparator that compares the desired, actual and predicted kinematic states, to generate sensorimotor errors used to drive motor learning.

The MT and MST brain regions are modulated by selective visual motion patterns (Duijnhouwer et al., 2013; Peuskens, Sunaert, Dupont, Van Hecke, & Orban, 2001; Tootell et al., 1995). Prior work has suggested the possible contributions of MT/MST to motion perception and various associated cognitive functions during goal-directed behaviors such as movement of the upper-extremity with a certain degree of complexity (Bisley & Pasternak, 2000; Born & Bradley, 2005; Ilg, 2008; Parker & Newsome, 1998; Tanaka & Saito, 1989). MT selectively responds to directional local motions and speeds of objects in the visual field (Born & Bradley, 2005; Duijnhouwer et al., 2013). MST receives visual input from MT, and is critical for higher visual motion processing since its neurons are tuned to particular directions of motion (Born & Bradley, 2005). In particular, MST is thought to contribute to solving the visual rotation problem (Duijnhouwer et al., 2013). Thus, in our model, the simulated MT/MST region provides selective visual motion information (motion direction, velocity) to the SPL and IPS so that these regions can perform their visuospatial transformations, as well as to the frontal and parietal MNS to perform action observation and imitation. As such, MT/MST enables a view-dependent representation of actions (Adelson & Movshon, 1982; Duijnhouwer et al., 2013; Kourtzi & Kanwisher, 2001; Tootell et al., 1995). In our model the MT/MST does not require any information from the STS since it is engaged earlier in the bottom-up visual stream processing.

The SPL and IPS regions are hypothesized to be heavily involved in the execution of mental rotation and spatial transformations. Much empirical evidence suggests that IPS/SPL play a critical role in the transformation between frames of reference, or viewpoints, since these areas encode spatially mapped representations (Buneo & Andersen, 2006; Culham & Kanwisher, 2001; Gogos et al., 2010; Grefkes & Fink, 2005; Zacks, 2008). It has also been suggested that the IPS/SPL areas may compute viewpoint transformations for self-intended movements (e.g., Buneo & Andersen, 2006; Zacks, 2008). These regions may embed mechanisms by which two frames

of reference can be merged through a particular relation resulting from associations learned through experience in observing rotational motion (Zacks, 2008). Thus, we postulate that these regions are involved in visuospatial transformations to subserve the MNS during learning and should enable view-independent representations.

The IFG of the MNS is activated during action observation (Buccino et al., 2001; Decety et al., 1997; Gallese et al., 1996; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Jeannerod, 2001; Rizzolatti et al., 2001). For instance, grasp observation activates the inferior frontal cortex, the supplementary motor area and the dorsal premotor cortex (Grafton et al., 1996). It has been proposed that the STS-IPL-IFG pathway acts as an inverse model that computes the motor commands for producing a given desired action (Miall, 2003). It has also been suggested that the motor/premotor regions may implement inverse models for self-intended reaching movements (Bullock, Grossberg, & Guenther, 1993; Gentili, Oh, Kregling, & Reggia, 2016; Gentili, Oh et al., 2015; Guenther & Micci-Barreca, 1997; Li, Padoa-Schioppa, & Bizzi, 2001; Molina-Vilaplana & Coronado, 2006; Molina-Vilaplana et al., 2007; Padoa-Schioppa, Li, & Bizzi, 2004; Srinivasa, Bhattacharyya, Sundareswara, Lee, & Grossberg, 2012; Tin & Poon, 2005). Thus, here we postulate that the IFG performs the inverse computation that maps a desired action such as an action to imitate, into the corresponding neural drive that allows imitating the observed action.

The IPL of the MNS becomes active during observation of an action such as grasping (Buccino et al., 2001; Decety et al., 1997;

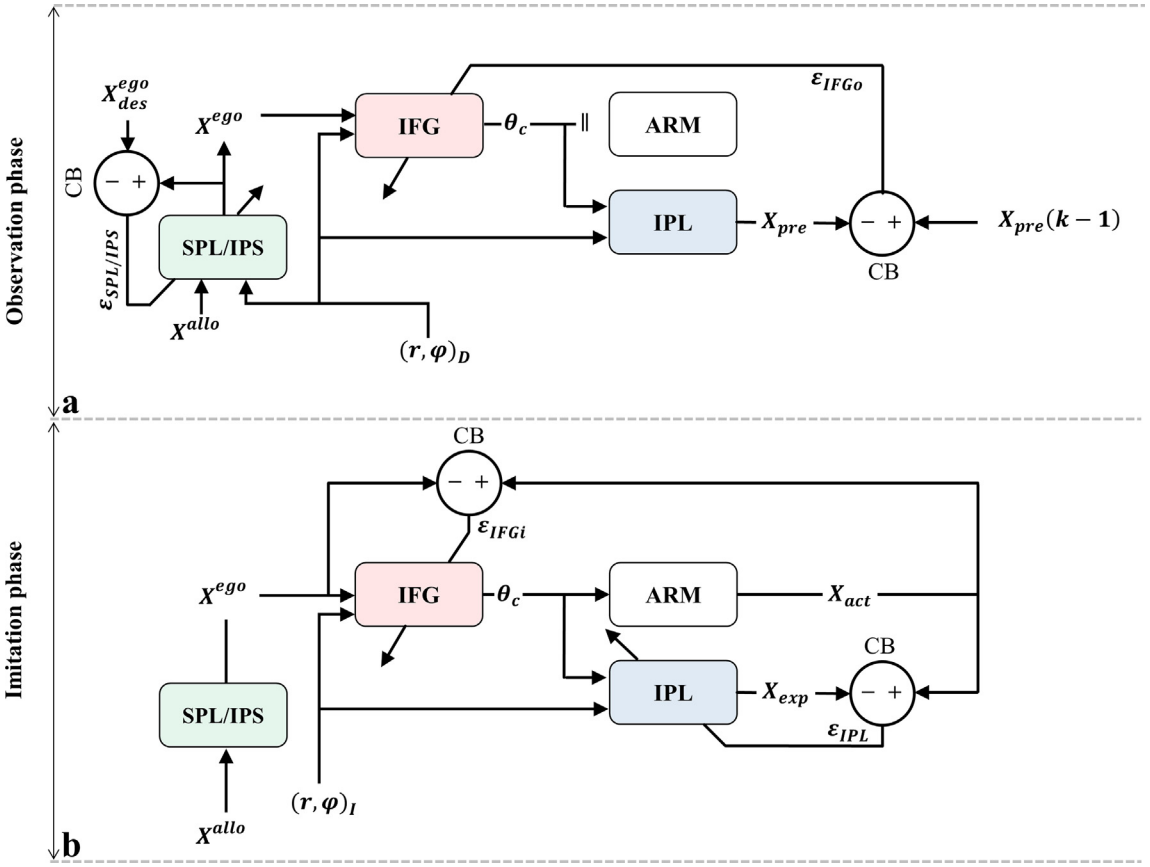


Fig. 2. Learning schemes of the parietal transformation network (SPL/IPS), frontal (IFG) and parietal (IPL) MNS during observation (a) and imitation (b). SPL/IPS, IFG and IPL learn via successive observation-imitation cycles. (a) First the SPL/IPS and IFG learn during the observation phase without the imitator executing any actual movements. The SPL/IPS network learns using an error signal ($\varepsilon_{SPL/IPS}$) provided by the cerebellum (CB) that acts as a comparator (indicated as circles with signs inside) that compares the desired (X^{ego}_{des}) and remapped spatial kinematics in the egocentric (X^{ego}) frame of reference. Simultaneously, a copy of the neural drive (θ_c, encoding shoulder and elbow joint kinematics) generated by the IFG is sent to IPL (forward model) which computes the predicted spatial kinematics ($X_{pre}(k)$) of the observed action. This predicted kinematics is covertly compared to that from the previous ($X_{pre}(k-1)$) trial by the CB which computes the error (ε_{IFGo}) to guide the learning of the IFG. (b) Then, the imitator imitates the action previously observed; actually carrying out the movement. The IFG network learns using an error (ε_{IFGi}) computed by the CB that compares the demonstrated action remapped into the allocentric space (action that aims to be imitated; X^{ego}) and the actual (X_{act}) spatial kinematics. Also, the IPL network learns using an error (ε_{IPL}) computed by the CB that compares the predicted (X_{pre}) and actual (X_{act}) spatial kinematics. During the imitation phase the SPL/IPS is switched to the identity function and thus the allocentric and egocentric frames of reference coincide perfectly ($X^{allo} = X^{ego}$). (r, φ)$_I$ and (r, φ)$_D$ are the pairs formed by the distance and orientation between imitator and demonstrator's end effector from the imitator's perspective, respectively. X^{allo}, X^{ego} and X^{ego}_{des}, are the observed action represented in the allocentric space, estimation of the observed action remapped into the egocentric space, and desired action to imitate (i.e., observed action remapped in the egocentric space). For clarity the rPFC and STS are not shown.

Grafton et al., 1996; Jeannerod, 2001; Rizzolatti et al., 2001). During observation of object-related actions there is also activation of the posterior parietal lobe (Buccino et al., 2001; Grafton et al., 1996). It has also been suggested that the STS-IPL-IFG pathway employs a copy of efference of the neural drive (from the IFG) to predict the corresponding sensorimotor consequences (i.e., forward computation) (Miall, 2003). The posterior parietal regions along with the IPL may also embed forward models that predict sensorimotor consequences during both overt and covert self-intended movement execution (e.g., Blakemore & Sirigu, 2003; Crevecoeur & Scott, 2014; Crevecoeur et al., 2011; Gentili, Oh et al., 2015; Gentili & Papaxanthis, 2015; Gentili et al., 2016; Gueugneau, Schweighofer, & Papaxanthis, 2015; Sirigu et al., 1996; Wolpert, Goodbody, & Husain, 1998).

2.3. Modeling learning by observation and imitation: a two-phase process

In our model shown in Fig. 1(b), the learning processes are inspired by the idea that, in humans, typically an imitator observes an action that is performed by a demonstrator, and subsequently attempts to reproduce it. Thus, there is first a learning by observation phase that corresponds to a covert learning process (i.e., no actual action) before an overt action is actually performed. Then, learning by imitation occurs involving the use of peripheral feedback as the movement is actually executed. This cycle is continuously repeated until the imitator can successfully imitate the action. When considering an ecologically valid situation, a human imitator can imitate an observed action regardless of differences in body sizes, spatial positions, or viewpoints between the demonstrator and the imitator. Such robustness in visuospatial information processing is critical for action observation and imitation. Our model learns actions in a similar manner by continuously executing sequences that combine observation and imitation phases independently of the differences in anthropometry, distance, and viewpoint between the demonstrator and the imitator.

During the *observation phase*, the imitator observes the action performed by the demonstrator in the allocentric frame of reference. The visual input that encodes observed actions is sent to the MT/MST for processing the motion direction and velocity, which are clearly view-dependent. This view-dependent information is then sent to the MNS network (IFG, IPL, and STS) and also to the SPL/IPS, which transforms it into a visuospatial representation in the imitator's view-independent egocentric frame of reference. Subsequently, the STS, which responds only to familiar motion of specific body parts (e.g., arm, hand), determines whether the observed action is currently in the imitator's motor repertoire or not. Specifically, when the imitator observes a reaching action, the following information is extracted from that observed action: the unspecific effector (i.e., arm without specification of right or left), type of action (i.e., reaching), and the associated trajectory (i.e., a reaching vector from an initial to a final position). Thus, if an observed reaching action is recognized in the STS (i.e., if the action is in the imitator's motor repertoire), the rPFC would simply trigger its imitation. If the action is not recognized, the rPFC triggers the observational learning mode, in which, the STS will encode the corresponding information mentioned above (i.e., effector, type and trajectory). In particular, when learning by observation is triggered, even though no action is overtly executed, the IFG performs the visual-to-motor inverse computation to generate the neural drive that would be used if the action was overtly executed but this execution is then inhibited. However, an efference copy is still available and is sent to the IPL which predicts the sensory consequences of the corresponding action (i.e., forward computation). The CB computes the prediction error that is conveyed to the IFG to adjust its internal model. At this point, the STS assesses the similarity between the imitator's sensory predictions and the corresponding observed action (Iacoboni, 2005). If the similarity is low due to a large error, the learning by observation continues until the error is acceptable. Once the observed action is encoded fairly well it can be imitated successfully (Fig. 2(a)).

Once the learning by observation phase is complete, learning during the *imitation phase* is performed in which action execution is triggered by the rPFC (Decety et al., 1997). During learning by imitation, the overall processes in the IFG, IPL, STS, CB, SPL/IPS and MT/MST are similar to those described in the observation phase. The only difference is that the neural drive is no longer inhibited and is thus sent to the musculoskeletal system via the premotor/motor regions to imitate the action. The imitator then observes its own action via visual feedback to adapt it further if needed. In this case, the imitator uses the identity visuospatial transformation since the self-action is encoded in its own egocentric frame of reference. The visual feedback updates both the inverse and forward model in the IFG and IPL, respectively. Once the imitation phase is complete, the next observation-imitation cycle starts and the process is repeated until the action is learned (Fig. 2(b)).

2.4. Model implementation

The demonstrated/imitated action is modeled as a reaching-grasping task where for the sake of simplicity the hand is aligned with the forearm and the control of finger grasp is not explicitly modelled (similarly to prior relevant; Sauser & Billard 2005a, 2005b, 2007). The demonstrator/imitator's hand/arm is modeled as a simple two links effector with 2 degree of freedom which moves in a 2D workspace. For simplicity the dynamics are ignored in particular since viewpoint transformations which relate to geometrical features of the hand/arm system are examined. Similar relevant computational work has employed a comparable approach where the dynamics were not modeled (Sauser & Billard 2005a, 2005b, 2007; Fleischer et al., 2013). It must be noted that while the reaching-grasping task employed in this work may appear simple, the underlying neural processes engaged during observation and imitation which are here captured with three interacting adaptive neural networks are very non-trivial (Sauser & Billard, 2005a).

Our neural model is implemented based on the conceptual and computational principles described above[1]. The rPFC, STS, CB, and

[1] The model was implemented and simulated using MATLAB® R2015a (MathWorks, Inc.) on Windows platform (Windows 7 Enterprise 64-bit platform on Optiplex 9020 Desktop, Dell, Inc.).

MT/MST are not implemented with artificial neural networks but instead with numerical and conditional statements. The SPL/IPS, IFG, and IPL are implemented by means of radial basis function (RBF) neural networks that learn their respective hypothesized functions at a population level, which are the visuospatial transformations, the inverse and forward computations, respectively. Our aim was not to model the cortical circuits as a biologically-realistic model of spiking neurons but instead to capture the main functionalities of the neural regions considered here. RBF neural networks modeling was selected since it is a well-established approach that was previously employed to model view-based processes in the context of 2D visual processing (e.g., Deneve & Pouget, 2003) as well as the MNS (e.g., Fleischer et al., 2013) and more generally in sensorimotor modeling of reaching and grasping actions (e.g., Han et al., 2008; Molina-Vilaplana & Coronado, 2006; Molina-Vilaplana et al., 2007; Porrill & Dean, 2007). Each neural network was trained by continuously repeating the two-phase observation-imitation learning cycle described above (Fig. 2). First we present several neural components of our architecture that were not implemented with neural networks since they were not central to our focus which was the interactions between the visuospatial transformation mechanisms processing the frame of references re-mapping and the MNS. Then the neural regions implemented with neural networks are presented.

The rPFC was implemented with a simple conditional statement that switches between the phase of learning by observation and learning by imitation. Specifically, when the imitator learned an observed action, a state variable representing the rPFC was set to the observational learning mode, whereas this variable was set to the imitation execution mode when the imitator reproduced the observed action. The STS was implemented as typical associative arrays through hash tables, in which the observed action was indexed by using a pair of body segment and the template of the observed action trajectories as keys to model observational learning.

The CB was implemented as a simple comparator to compute sensorimotor errors based on comparisons between desired, actual and expected kinematics that are then employed to guide the learning of SPL/IPS, IPL and IFG neural networks to acquire their respective mappings (see Fig. 2).

The MT/MST extracts the direction of the observed action represented by a 2D vector (Fig. 1(c)). The first component of these input vectors is the distance r between the imitator's center point and the demonstrator's end-effector at each time t. The second component is the relative angle φ from the imitator's viewpoint to the end-effector at time t. Thus, this vector can be written as:

$$a(t;0 \leqslant t \leqslant n) = \{(r_0, \varphi_0), \ ...,(r_t, \varphi_t), \ ...,(r_n, \varphi_n)\} \tag{1}$$

where a is an observed action from time 0 to time n. As such, the view-dependent representation of the action is described by a 2D vector which encodes the distance (r) and orientation (φ) of the demonstrator's effector (i.e., allocentric) from the imitator's viewpoint. When the imitator observes its own action from an initial to a final position, in the same way, the view-dependent information is also extracted (r, φ), which is this time encoded in its own egocentric frame of reference. Thus, the MT/MST provides to the MNS and the SPL/IPS the view-dependent representation of the observed action (Fig. 1(b)).

The SPL/IPS implement a visuospatial transformation (named f_{VST}) which combines translation, scaling, clockwise and counterclockwise rotation by employing a RBF network (Broomhead & Lowe, 1988a, 1988b). The use of modeling clockwise and counterclockwise rotation as well as scaling and translation mechanisms is consistent with prior neurophysiological and behavioral human studies which revealed that those regions have a critical role in such visuospatial/visuomotor transformations. Specifically, the functional roles of SPL/IPS in these visuospatial transformations (i.e., clockwise, counterclockwise rotations, translation, and scaling) have been evidenced using various techniques including visual field tests as well as fMRI (Burton, Wagner, Lim, & Levy, 1992; Cohen et al., 1996; Grefkes, Ritzl, Zilles & Fink, 2004; Kadmon Harpaz, Flash, & Dinstein, 2014). The translation, scaling and rotation are defined as follows: i) *translation*: mechanism for the imitator to refer to the demonstrator's actions in the same position by shifting the origin of the imitator's frame of reference; ii) *scaling*: allows the imitator interpreting the observed actions to change the ratio and shape of the body of the demonstrator, and iii) *rotation*: allows the imitator to face the same direction as the demonstrator by rotating the orientation of the axial frame of the imitator (Frank, 1998; Lopes & Santos-Victor, 2005). In particular, the rotation angle around the z axis θ_z is the angular displacement for the mental rotation from the demonstrator to the imitator's viewpoint:

$$\theta_z = \theta_I - \theta_D \tag{2}$$

where the rotation is counterclockwise if $0^\circ < \theta_z \leq 180$, and clockwise if $-180^\circ < \theta_z < 0^\circ$, respectively (Fig. 1(d)).

The SPL/IPS network receives as inputs the visuospatial information that encodes i) the kinematics of the demonstrator expressed in the allocentric frame of reference (i.e., $X^{allo} = [x^{allo}, y^{allo}]$ inputs to SPL/IPS in Fig. 1(b)) and ii) the imitator-demonstrator viewpoint (i.e., r and φ inputs to SPL/IPS in Fig. 1(b)) provided by the MT/MST. The network outputs are the remapped visuospatial information (i.e., $X^{ego} = [x^{ego}, y^{ego}]$ output of SPL/IPS in Fig. 1(b)) in the imitator's frame of reference. This network is trained using supervised learning in which the targets were provided by combining the three transformations mentioned above (translation, scaling, clockwise/counterclockwise rotation) which were implemented with neural networks previously learned (for details about the implementation of these four transformation networks, see sections S1 and S2 of the online supplementary material). It must be noted that although the primitives were learned first, a successful combined learning of the IFG, IPL (see below) and SPL/IPS networks was not guaranteed and as such this preliminary learning was necessary but not sufficient for acquiring the various sensorimotor transformations learned by the proposed neural model. Such an approach allowed the SPL/IPS network to learn any configuration of the positions, frames of reference, and viewpoints between the demonstrator and the imitator to remap the visuospatial information related to an observed action from the demonstrator's allocentric into the imitator's egocentric frame of reference. This mapping is the identity during action imitation since the imitator self-observes its own limb motion.

Specifically, this RBF network was trained based on the supervised learning technique proposed by Orr (1998). Forward subset selection (Miller, 1984) using orthogonal least square (OLS) (Chen, Cowan, & Grant, 1991) was employed to determine an optimal subset of the available centers. Moreover, generalized cross-validation (GCV) was used to define model selection criterion with

additional parameters that hold temporarily the selection process to estimate how the trained network performs for unknown inputs (Golub, Heath, & Wahba, 1979). This learning procedure was initially performed on coarsely spaced RBFs which, through progressive refinement, identify the best RBF width. Learning stops adding further RBFs when the decreasing ratio of the GCV was less than 1.0×10^{-4} for at least two iterations. The same learning stop criterion was applied for the training of the other neural networks. Finally, backward elimination is employed to selectively remove less significant RBFs. As such, this network had 4 input neurons (x^{allo}, y^{allo}, r, φ); one single hidden layer which was composed of 47 RBF neurons after learning (23 and 24 for the x and y dimension, respectively) and 2 output neurons (x^{ego}, y^{ego}) (for more details about the SPL/IPS implementation, see sections S1 and S3 of the online supplementary material).

The IFG was also implemented by means of a RBF network and encoded an internal inverse model to guide the imitator's arm and hand to the target object. This neural network learns the inverse computation of the forward kinematics of the upper extremity which is described by:

$$\begin{aligned} x &= l_1\cos\theta_1 + l_2\cos(\theta_1 + \theta_2) \\ y &= l_1\sin\theta_1 + l_2\sin(\theta_1 + \theta_2) \end{aligned} \tag{3}$$

where x and y designate the end-effector position of the imitator, l_1 and l_2 are the length of the upper-arm and forearm, respectively, while θ_1 and θ_2 are the shoulder and elbow joint angles, respectively. Therefore, the inverse mapping (f_{INV}) computed by the IFG network can be described as[2]:

$$(\theta_1, \theta_2) = f_{INV}(x, y, r, \varphi) \tag{4}$$

where x and y designate the desired spatial end-effector position that the imitator aims to reach (in its own frame of reference), where the angle φ is the relative angle from the imitator's viewpoint to the demonstrator's workspace where the demonstrator's end-effector is placed, and r is the distance to this point from the imitator's center point (see Eq. (1) and Fig. 1(c)). In addition, the angular velocity of the end-effector with respect to the imitator's viewpoint (θ_I), which is used by the IFG network, was computed to represent the visual motion information that encodes the perspective by employing the following equations.

$$\begin{cases} \dot{\varphi} & = \frac{1}{r}\|v_\perp(\dot{x}, \dot{y})\| sin\,\varphi \\ v_\perp(\dot{x}, \dot{y}) & = \begin{pmatrix} \cos\theta_I & \sin\theta_I \\ -\sin\theta_I & \cos\theta_I \end{pmatrix} \begin{pmatrix} \partial x/\partial\theta_1 & \partial x/\partial\theta_2 \\ \partial y/\partial\theta_1 & \partial y/\partial\theta_2 \end{pmatrix} \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} \end{cases} \tag{5}$$

where $v_\perp$ is the perpendicular component of the motion (angular velocity) with respect to φ, and $\|\cdot\|$ denotes the L^2-norm (for additional details about the anthropometric data see section S4 of the online Supplementary Material).

The training of the inverse model during the observation and the imitation stages was guided by the predicted and actual end-effector position, respectively (see Fig. 2(a) and (b)). As such, it is reasonable to consider that the weight of the prediction of the sensory consequences is more important during the learning of the inverse model in the observation phase since the imitator mentally simulates the observed action (Case et al., 2015; Hétu et al., 2013). Conversely, during the learning that takes place in the imitation phase, the visual feedback of the self-executed action has greater influence on the learning of the inverse model. Therefore, to account for these differences, Eq. (4) was re-written as:

$$(\theta_1, \theta_2) = f_{INV}(\widetilde{x}, \widetilde{y}, r, \varphi) \tag{6}$$

where the input vector $(\widetilde{x}, \widetilde{y})$ is the end-effector position represented as follows:

$$\begin{pmatrix} \widetilde{x}, \widetilde{y} \end{pmatrix} = \begin{cases} (1-\delta)(x, y) + \delta(x_{pre}, y_{pre}), & \text{in learning by observation} \\ \delta(x, y) + (1-\delta)(x_{pre}, y_{pre}), & \text{in learning by imitation} \end{cases} \tag{7}$$

The weighting factor δ ($0.5 \leqslant \delta \leqslant 1$) ensures that the predicted end-effector positions are preponderant in the absence of peripheral feedback during learning in the observation phase, whereas the role of the visual feedback end-effector positions is emphasized during the learning in the imitation phase where the actual action is executed. The predicted spatial end-effector position (x_{pre}, y_{pre}) is generated by the IPL.

As such, the IFG neural network receives visuospatial information inputs composed of a component independent of the perspective (i.e., $X^{ego} = [x^{ego}, y^{ego}]$ inputs to IFG in Fig. 1(b)) as well as a dependent one (i.e., r and φ inputs to IFG in Fig. 1(b)), each being provided by the SPL/IPS and MT/MST, respectively (X^{ego} is the observed spatial kinematics remapped into the egocentric imitator's frame of reference). The output of the IFG network is the motor command ($\theta_c = [\theta_1, \theta_2]$ outputs from IFG in Fig. 1(b)) encoded as the joint angles to produce the action previously observed. Thus, the IFG network had 4 input neurons (x^{ego}, y^{ego}, r, φ); one single hidden layer which after learning was composed of 94 RBF neurons (46 and 48 for the shoulder and wrist joint, respectively) and 2 output neurons (θ_1, θ_2). The IFG neural network was trained with the same learning procedure described above as for the SPL/IPS network (Broomhead & Lowe, 1988a, 1988b; Orr, 1998). The weighting factor was initially set to a high value (e.g., 0.9) and gradually decreased over time during training, resulting in a prediction error scaled between 0.5 and 1 (for more details about the implementation of the IFG network, see sections S1 and S5 of the online Supplementary Material).

[2] It must be noted that r and φ represent inputs that allows simulation of the encoding of the view-based information in the frontal MNS.

Finally, the IPL was implemented with a RBF network in order to encode an internal forward model that computes the mapping (f_{FWD}) that can be described as[3]:

$$(x_{\text{pre}}, y_{\text{pre}}) = f_{FWD}(\theta_1, \theta_2, r, \varphi) \tag{8}$$

where x_{pre} and y_{pre} designate the predicted spatial end-effector position (in the imitator's own frame of reference), the angleφ is the relative angle from the imitator's viewpoint to the demonstrator's workspace where the demonstrator's end-effector is placed, and r is the distance to this point from the imitator's center point (see Eq. (1) and Fig. 1(c)). As such, the IPL network receives as inputs a copy of the neural commands (copy of efference) that encode the joint kinematics[4] (i.e., $\theta_c = [\theta_1, \theta_2]$ inputs to IPL in Fig. 1(b)) as well as visuospatial information input (i.e., r and φ inputs to IPL in Fig. 1(b)) that dependent of the viewpoint, each being provided by the IFG and MT/MST, respectively. The outputs of the IPL network are the predicted spatial end-effector positions ($X_{pre} = [x_{\text{pre}}, y_{\text{pre}}]$ outputs from IPL in Fig. 1(b)) that mimic the sensory consequence of the neural commands generated to produce covertly or actually the action previously observed. Thus, the IPL network had 4 input neurons (θ_1, θ_2, r, φ); one single hidden layer which, after learning was composed of 53 RBF neurons (27 and 26 for the x and y dimension, respectively) and 2 output neurons (x_{pre}, y_{pre}). Like for the SPL/IPS and the IFG networks, the IPL network was trained with the learning procedure previously described (Broomhead & Lowe, 1988a; Broomhead & Lowe, 1988b; Orr, 1998; for further details about the IPL network implementation, see sections S1 and S6 of the online Supplementary Material).

2.5. Learning by the SPL/IPS, IFG and IPL networks

Learning by the three neural networks SPL/IPS, IFG and IPL was systematically conducted through a two-phase observation-imitation cycle. This learning strategy is ecologically valid since it is reasonable to expect that an action will be observed first (i.e., learning by observation) before it is actually performed with the intent to reproduce it (i.e., learning by imitation). Generally, the same process needs to be repeated several times for complete acquisition of a skilled action. As such, the SPL/IPS and IFG are trained during a learning by observation phase (Fig. 2(a)), whereas both the IFG and IPL are trained during the action imitation stage (Fig. 2(b)). Moreover, contrary to various existing neural models of human motor behavior where each network is trained separately, here an online updating method is employed so that the three neural networks learn together. The online method updates all the neural weights sequentially (SPL/IPS followed by the IFG, followed by the IPL, and then IFG again) by employing the corresponding errors computed for each step. This is repeated until all three networks are completely trained. Such a learning strategy is more realistic than training separately the SPL/IPS, IPL and IFG networks since it simulates the interactions between these three neural components. The learning stage uses continuous iterations of two-phase learning via the observation-imitation cycle (for more details about the error signals that drive the learning of each network, see Fig. 2).

2.6. Model performance assessment

A simple geometrical model of the arm having two degrees of freedom was used to perform horizontal reaching task in a 2D workspace. The task was performed under nine conditions that are described in the Cartesian plane by a combination of three distances (near (1.50 m), middle (2.50 m), and far (3.50 m), along a depth axis) and three lateral positions (left (−45 cm), center (0 cm), and right (+45 cm)) which represent the demonstrator's relative position with respect to the position and viewpoint of the imitator (Fig. 1(e)). This was done by simulating a simple environment where the distances and lateral positions of the demonstrator were changed by modifying the Cartesian position of the demonstrator with respect to the imitator which remained at the same position defined as the coordinate of origin (0, 0) of this environment.

Moreover, for more realistic conditions, depth perception was implemented to allow the imitator to perceive a smaller size of the demonstrator when it is farther away. The imitator's depth perception ratios were 97%, 95%, and 93% of the demonstrator's original body size for the near, middle, and far conditions, respectively. For all conditions, the imitator performs the reaching action through one observation-imitation cycle while facing the demonstrator. Thus for each given condition, two tests are performed which correspond to observing and imitating the demonstrator's action. A vector-integration-to-endpoint (VITE) model is used to generate the demonstrated trajectories for each action (duration: 1.5 s; sampling rate: 128 Hz; Bullock, Bongers, Lankhorst, & Beek, 1999; Bullock & Grossberg, 1988).

After training the MNS model performance was assessed by considering: i) the learning curves produced by each network along with the associated magnitude of errors, and ii) both process (e.g., spatial and joint kinematics of the effector) and outcome (e.g., processing time during visuospatial transformation) measures as well as the distribution of the synaptic activity. First, the learning performance of the networks was assessed through their learning curves expressed as the changes in mean squared error (MSE) for the OLS and GCV methods for each condition. In addition, since the computation of the root-mean-square error (RMSE) is a usual metric to assess and compare human performance under various conditions, it was employed to assess the performance of the neural model (Buch, Young, & Contreras-Vidal, 2003; Gentili, Bradberry, Oh, Hatfield, & Contreras-Vidal, 2011; Gentili et al., 2015; Kagerer, Contreras-Vidal, & Stelmach, 1997). Namely, the networks' performance was assessed by computing the RMSE between the imitator's arm end-position in the imitator egocentric frame of reference (after remapping by the neural network) and the expected imitator arm's end-position which

[3] It must be noted that r and φ are used as inputs to simulate the encoding of the view-based information in the parietal MNS.

[4] Note that the joint kinematics are in essence encoded in the imitator's egocentric frame of reference and thus are independent of perspective.

would be performed if this remapping was perfect (i.e., the desired end-position to be reached obtained via the exact transformation). In particular, during the observation phase the RMSE was computed between these desired end-positions and the predicted imitator's arm end-positions (output of the internal forward model). During the imitation phase, the same computation was implemented the only difference was that the actual imitator's arm end-positions were employed instead of the predicted ones.

Second, the quality of action imitation was tested by assessing planar arm reaching movements where the demonstrator-imitator viewpoint was manipulated. Specifically, the kinematics including the displacement and velocity for both the end-effector and the two joints of the upper-extremity were analyzed. In order to test the neural model's visuospatial transformation capabilities, execution of mental rotations was performed under various conditions that consisted of manipulating the position of the demonstrator to change the viewpoint of the observed scene. Namely, such assessment included clockwise and counterclockwise rotations as well as scaling and translation transformations, although the assessment of the last two was emphasized less. In particular, mental rotation capabilities of the SPL/IPS were examined by measuring the relationship between the rotation angle and the response (or processing) time,[5] which was defined as the duration for the network to successfully transform the observed action from the allocentric into the imitator's egocentric frame of reference.

Lastly, the neural dynamics of the three networks was investigated by computing their local mean synaptic activity where positive and negative activity corresponded to excitatory and inhibitory synaptic activity, respectively. For these three neural networks, each representing a brain region, the mean synaptic activity during the action was computed assuming that each network node would reflect a subpopulation of neurons rather than a single neuron. As such, the obtained mean synaptic activity represented the regional mean neural population activity (or local-field potential) (David & Friston, 2003; Jansen & Rit, 1995). In addition, an index of dissimilarity was computed to compare the similarity between two matrices of synaptic activity using the covariance matrix distance which was proposed based on the correlation matrix. The covariance matrix distance metric ranges from 0 to 1, indicating that two matrices of synaptic activity are completely different or very similar, respectively (Herdin, Czink, Ozcelik & Bonek, 2005) (for further details about the computation of the synaptic activity and of the dissimilarity index, see sections S7 and S8 of the online Supplementary Material, respectively).

3. Results

Overall, the findings revealed that after learning, the SPL/IPS, the IFG and the IPL successfully encoded the visuospatial transformation, and the inverse and forward models, respectively. The acquisition of sensorimotor transformations allowed the neural model to perform action observation/imitation independently of the differences in distance, anthropometry and viewpoints between the demonstrator and the imitator. Specifically the results include, i) the imitator was able to accurately imitate the action previously shown by the demonstrator while generating kinematics similar to those observed in humans; ii) the model predicts that during observation/imitation the response time to perform mental rotation from various differing viewpoints processed by the SPL/IPS regions was independent of the direction (i.e., clockwise or counterclockwise) of a given angle but dependent on the amount of rotation observed; iii) the network's synaptic activity in the frontal and parietal MNS was similar during both observation and imitation (under the same conditions) while also revealing in the frontal MNS the emergence of both view-independent and view-dependent representations of observed actions and iv) the model predicts that IPL includes essentially view-independent neurons whereas SPL/IPS embeds both view-independent and view-dependent neurons. The findings (i, iii) are in agreement with behavioral and neurophysiological studies, respectively while the results (ii, iv) provide behavioral and neurophysiological predictions that could be empirically tested. For brevity, only three conditions (left-middle, center-middle, and right-middle) are described in detail below, but very similar findings were obtained for the other six conditions.

3.1. Performance throughout learning

After learning, the three neural network components (SPL/IPS, IFG and IPL) successfully encoded their respective sensorimotor mappings (Fig. 3). Interestingly, the IPL tended to be acquired faster than the IFG (compare light gray and dark gray curves, respectively) independently of the testing conditions. Overall the MSEs were small ($\leq 1.63 \times 10^{-2}$) across conditions. During the early learning stages an initial peak was occasionally observed (e.g., see the SPL/IPS learning curve in the middle panel of Fig. 3). This was likely due to two reasons. First, the learning method employed here (OLS, GCV) adds a new hidden node during each iteration, resulting in an adjustment of all the network parameters. Second, the concomitant learning of the three networks can lead to the propagation of erroneous output from a given network to the others during the early stage of learning, resulting in the observed initial increase of error in the learning curve. As the three networks learn further such a peak vanishes.

3.2. Performance after learning

Independently of differences in orientation, distance and anthropometry between the imitator and demonstrator, the neural model was able to reproduce accurately the action (i.e., typical spatial trajectories, displacements and velocity profiles) previously observed (Fig. 4).

[5] Under the assumption that the processing time is linearly proportional to the behavioral reaction time, this metric can be related to the reaction time measured at the behavioral level.

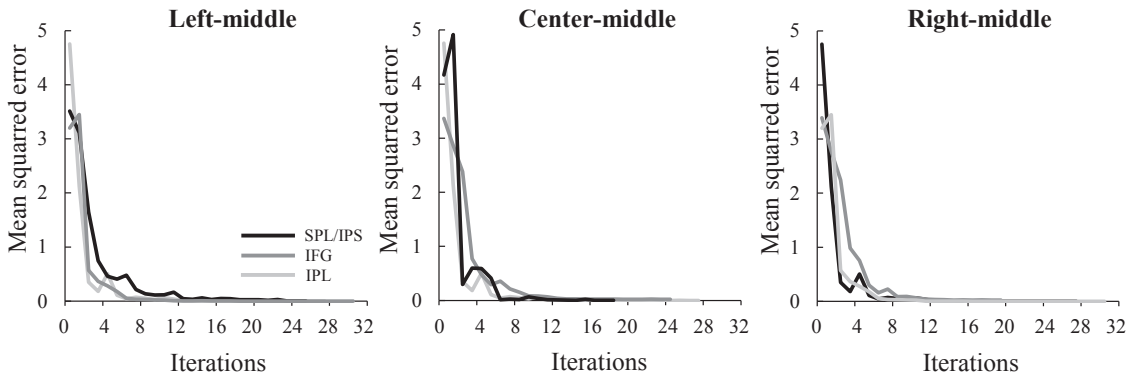


Fig. 3. Learning curve expressed as the average of OLS and GVC values for the three neural networks (SPL/IPS, IFG and IPL) during learning with an online updating method. The learning curves for SPL/IPS, IFG and IPL are represented in black, dark-gray and light-gray, respectively. The results for the left-middle, center-middle and right-middle imitator position are depicted in the left, center and right panels.

3.2.1. Behavioral performance

During both action observation and imitation, the mean values and standard deviations of the RMSE were obtained in the left-middle, center-middle, and right-middle conditions. Overall, the RMSE were similar in action observation and imitation across testing conditions (the difference of mean RMSE for observation and imitation was ≤ 1.1 mm which is $\leq 0.39\%$ of the length of the imitator's limb; all conditions considered).

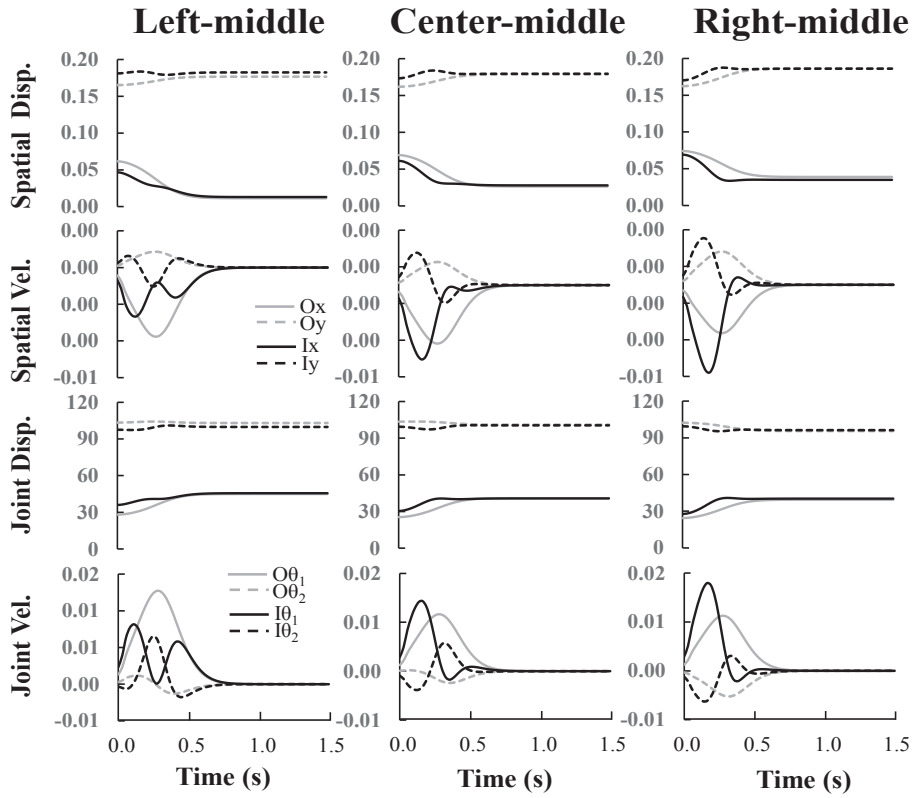


Fig. 4. Demonstrated and imitated hand/arm kinematics executed from three different viewpoints (imitator's left, center and right). Displacement and velocity profiles of the demonstrated and imitated actions to targets in both spatial and joint coordinates. The gray and black lines represent the predicted and actual kinematics during action observation and imitation, respectively. For the spatial displacement and velocity, the solid and dashed lines represent the x and y dimension, respectively. For the joint displacement and velocity, the solid and dashed lines represent the shoulder and elbow joint, respectively. Ox and Oy: spatial kinematics during action observation on the x and y dimension, respectively. Ix and Iy: spatial kinematics during action imitation on the x and y dimension, respectively. $O\theta_1$ and $O\theta_2$ joint kinematics for the joint θ_1 and θ_2 during action observation. $I\theta_1$ and $I\theta_2$ joint kinematics for the joint θ_1 and θ_2 during action imitation. The spatial displacements are in meter while the joint displacements are in degree.

Such low RMSE values resulted in similar spatial and joint kinematics for both observed and imitated actions while also being comparable to those previously observed in humans. Namely, after learning the obtained typical kinematics generally revealed sigmoid shape displacements and bell-shaped symmetrical velocity profiles as experimentally observed in human (Abend, Bizzi, & Morasso, 1982; Gentili, Cahouet, & Papaxanthis, 2007; Gordon, Ghilardi, & Ghez, 1994; Molina-Vilaplana & Coronado, 2006; Morasso, 1981).

However, occasionally displacements and velocity profiles generated by the imitator included small additional peaks or at least some modulations. Those are due to the visual feedback used by the imitator to regulate its imitation performance to reproduce as well as possible the action previously observed. Such a feature was not preprogrammed *per se* but was an emergent property resulting from the two-phase observation-imitation cycle on which the neural model was trained (Fig. 4).

SPL/IPS performance was assessed by computing the response time required to complete the visuospatial transformations between various viewpoints during action observation. Since this transformation included rotation, translation, and scaling, the response times reflect the duration that the SPL/IPS needed to complete these operations. The model predicts that the normalized response time to transform the observed action was: i) independent of the direction of rotation (i.e., clockwise or counterclockwise) for a given angle, and ii) monotonically increasing in a linear fashion as a function of the amount of the rotation angle at which the action is perceived. For the special case of 0°, the SPL/IPS performs only translation and scaling, and for the other special case of 180°, it uses counterclockwise rotation (Fig. 5).

3.2.2. Network synaptic activity change

The neural activity of the SPL/IPS, IPL and IFG networks was computed under various conditions of distances and viewpoints. As hypothesized, the neural activities of each network were comparable (under the same conditions) for both observation and imitation learning independently of the distance and viewpoint from which the reaching movement was observed and subsequently imitated.

The findings reveal that the neural activity of both excitatory and inhibitory synapses were very similar to visual inspection for both action observation and imitation (Fig. 6(a)–(b)). The synaptic patterns of the IPL were the same regardless of the testing conditions whereas those generated by the SPL/IPS and IFG had slightly different patterns in each condition. More quantitatively, the results revealed that the dissimilarity of the average synaptic activity of each neural network were comparable during action observation and imitation across the assessment conditions. Such similar patterns of mean synaptic activity between action observation and imitation is important since it is consistent with the well-established observation that the MNS activity is similar during observation and execution of the same action (Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011; Gazzola & Keysers, 2009; Grèzes, Armony, Rowe, & Passingham, 2003; Turella, Erb, Grodd, & Castiello, 2009). Although the IFG reveals slight differences in the mean synaptic activity, such a discrepancy is still very small ($\leq 2.19 \times 10^{-17}$ Arbitrary Unit (AU)).

Thus, overall the results revealed that all three brain regions are similarly activated during action observation and imitation (Fig. 6(c)). Moreover, on average the IFG revealed the presence of view-dependent (87%) and view-independent (13%) neurons that are reminiscent of view-based mirror neuron responses found in primates since their responses were dependent and independent of the perspective from which the action was perceived during action observation/imitation, respectively (Caggiano et al., 2011;

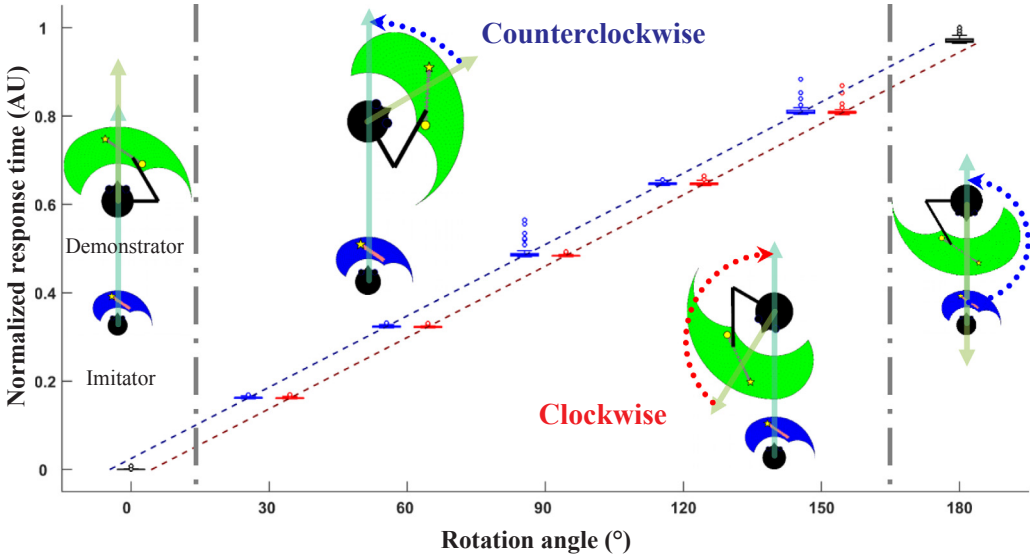


Fig. 5. Response time as a function of the rotation angle when the demonstrator and imitator are aligned on the vertical axis in the middle position. The rotation angle ranges from 0° (both demonstrator and imitator have the same orientation) to 180° (both demonstrator and imitator face each other). For a given angle between 0° and 180°, the simulated normalized response time was obtained for both the clockwise (red) and counterclockwise (blue) rotations. The mean and the standard deviations were calculated on 25 trials for each case. AU: Arbitrary Unit; (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

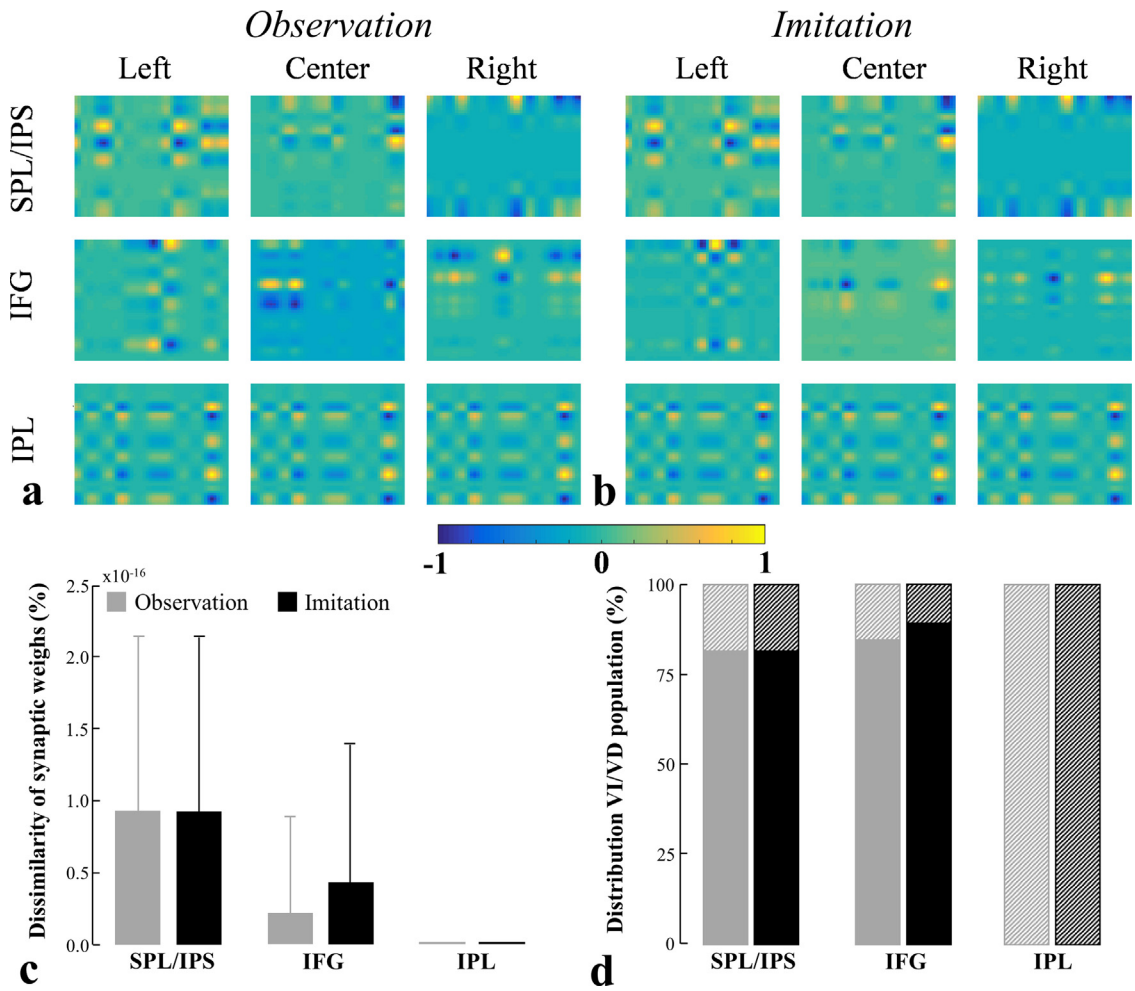


Fig. 6. (a–b) Mean synaptic activity for the SPL/IPS, IFG and IPL neural networks during action observation (a) and imitation (b). (c) Index of dissimilarity to assess more quantitatively the discrepancy of the mean synaptic activity between observation and imitation for the three neural networks. (d) Proportion of view-dependent (solid color portion of the bar) and view-independent (non-solid color portion of the bar) neural populations based on the mean synaptic activity patterns for the three neural networks. In panels (c–d) the gray and black fill represent action observation and imitation, respectively. For visualization, the mean synaptic activity of the RBF neurons are depicted in a 2D space (which represent the X × Y Cartesian space for both the SPL/IPS and IPL networks and represent the $\theta_1 \times \theta_2$ joint space for the IFG where θ_1 and θ_2 are the shoulder and elbow joints). The mean synaptic activity is normalized to be scaled from −1 (blue) to 1 (yellow) (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Caggiano et al., 2015). Importantly, the model predicts that the IPL would be mainly view-independent (i.e., 100%) while SPL/IPS would include both view-dependent (82.15%) and view-independent (17.85%) (Fig. 6(d)). Similar results were obtained for the six other experimental conditions.

4. Discussion

We have studied an adaptive neural architecture inspired by the organization and dynamics of the fronto-parietal network that learns to mimic the functional roles of the i) frontal MNS (IFG; inverse computation), ii) parietal MNS (IPL; sensorimotor predictions), and iii) parietal networks IPS/SPL. These latter brain regions support the observation and imitation functionalities of the MNS by performing visuospatial transformations such as rotations. Our neural model exhibited robust and accurate observation/imitation learning capabilities that involved reaching-grasping actions, and did so with similar hand/arm kinematics to those generated by the demonstrator, regardless of the demonstrator-imitator hand/arm viewpoint, distance or anthropometry. Critically, our model predicts that response time increases linearly (i.e., duration of the transformation in SPL/IPS) as the rotation angle of the observed action increases and that those changes in response time are similar when this rotation angle has a counterclockwise or a clockwise direction. This is an important emergent property of our neural model which could be experimentally tested by asking individuals to imitate reaching movements from various perspectives and measuring the response/reaction time.

Moreover, the synaptic patterns were similar in the frontal and parietal MNS during both action observation and imitation along with the emergence of both view-independent and view-dependent fronto-parietal network neurons in the macaque monkeys that are reminiscent of the actual view-based mirror neurons. This was not preprogrammed *per se* but is also an emergent feature of our neural model which is critical since it is consistent with i) the well-established notion that MNS activity is similar during observation and execution of the same action (Arnstein et al., 2011; Gazzola & Keysers, 2009; Grèzes et al., 2003; Turella et al., 2009) and ii) the identification of view-based mirror neurons in primates (Caggiano et al., 2011). More importantly, beyond replicating the ratio between view-independent and view-dependent neurons in IFG, the neural model predicts that IPL and SPL/IPS would mainly include view-independent and view-dependent neurons, respectively. This could also be examined by assessing the neural activity in these regions during action observation/imitation.

Overall, our neural model captures fronto-parietal computational mechanisms that learn to encode both sensorimotor and view-based internal representations. These representations enable visuomotor processing and serve as the basis for generating accurate neural commands and sensory estimates while imitating relatively fine motor skills where the observed arm and hand during the reach-to-grasp action have different geometrical features (e.g., orientations and size). Our model thus reproduces major behavioral and neurophysiological findings from prior relevant studies as well as generates behavioral and neurophysiological testable predictions in an observation/imitation context.

4.1. Visuospatial transformation as a support system for the fronto-parietal MNS

The visuospatial transformation processes involved during motor actions have been postulated to be implemented in the SPL/IPS based on past empirical studies (Andersen, 1987; Buneo & Andersen, 2006; Grefkes & Fink, 2005; Hesse, Sparing, & Fink, 2009; Oosterhof et al., 2012; Rizzolatti et al., 1998; Shmuelof & Zohary, 2008; Zacks, 2008). Our work here expands this notion to the MNS context by hypothesizing that the SPL/IPS performs an allocentric-to-egocentric transformation of observed actions, and sends these egocentrically-represented actions to the frontal MNS to prepare the corresponding neural commands (i.e., visual-to-motor transformation). Therefore, our neural model suggests that the IPS/SPL regions would be a critical part of the computational mechanisms described below that enable robust action observation and imitation. First, during action observation, the IPS/SPL regions would remap the visual inputs encoding a given observed action from an allocentric into an egocentric frame of reference. Once such a transformation is complete, an action that was initially observed from different perspectives is now encoded as the same action with a unique spatial orientation in the egocentric frame of reference of the imitator, independently of the viewpoint, anthropometry and distance between the imitator and demonstrator. Then, such a remapped action is provided to the IFG where it is encoded as the desired action to be overtly or covertly executed by computing the corresponding neural drives, regardless of the imitator-demonstrator visuospatial configurations. If the action is just observed, such a desired action is covertly executed by blocking the neural drive, whereas a copy of efference can still be sent to the IPL (forward model) to predict the sensory consequences that in turn are sent back to the IFG (inverse model) to refine the neural commands (i.e., simulation of action via fronto-parietal loops) (Crevecoeur, Thonnard, & Lefèvre, 2009; Gueugneau et al., 2015; Molina-Vilaplana & Coronado, 2006; Molina-Vilaplana et al., 2007). If the action is imitated, the entire process is the same, the only difference being that the neural drive is released to produce the actual action (i.e., overt execution). Note that the visuospatial transformation is part of sequential computational mechanisms which first involves the MT/MST region, which sends its computations to the parietal regions SPL/IPS and IPL, both in turn sending their outputs (remapping and sensorimotor predictions, respectively) to the IFG. The latter contributes to generating the neural drive. Interestingly, such sequential computational mechanisms are in agreement with the sequential engagement of similar cortical regions identified by Nishitani and colleagues during action observation and imitation. These also reinforce the biological relevance of this neural model (Nishitani & Hari, 2000, 2002).

As such, the overall computational mechanisms, and in particular those underlying the parietal visuospatial transformation, provide a coherent process whose components are similarly engaged during both action observation and imitation. As a result, the imitator can mimic the same action previously observed form various perspectives while producing similar kinematics. Occasionally some of these imitated kinematics reveal fairly slight discrepancies with observed ones since the imitator can regulate via visual feedback the overt imitation of the action previously observed. This was a characteristic not explicitly preprogrammed into our model that emerged as a result of the observation-imitation cycle employed to train it. During training the IFG and IPL acquired successfully their respective mappings to generate the imitated reaching action. However, their interactions can still result in the accumulation of small errors due to their respective residual learning errors. Thus, when generating the neural drive (feedforward control) such IFG-IPL interactions may produce occasional small discrepancies between portions of the demonstrated and imitated action which are naturally corrected using peripheral feedback (i.e., feedback control). While providing additional training and/or changing its implementation may minimize such discrepancies, here we aimed to simulate a model combining observation and imitation through multiple neural networks and not necessarily to obtain an extremely high accuracy performance such as that which would be required for engineering applications.

The parietal visuospatial transformation processes in our neural architecture were implemented in the SPL/IPS which, after learning, was able to perform clockwise and counterclockwise rotations as well as translation and scaling. Specifically, during action observation/imitation our model predicts i) similar response time when performing both counterclockwise and clockwise rotations, and ii) a linear increase in such response time as the magnitude of the rotation angle (of the observed action) increases. These findings suggest that complex mental rotations require more processing time before reaching the MNS and that the SPL/IPS processing time increases linearly with respect to the view angle. This is in agreement with the general framework of the neural simulation of action theory, and extends such mental rotation mechanisms from the cognitive to the cognitive-motor domain (Jeannerod, 2001). While

similar behavioral results were obtained during purely cognitive tasks such as mental rotations (Bock & Dalecki, 2015; Dalecki, Hoffmann, & Bock, 2012; Schwabe, Lenggenhager, & Blanke, 2009), as far as we know these results were not extended to a context of action observation/imitation. Our model predicts that similar neural processes would be employed during action observation and imitation under various viewpoints potentially extending those mental rotation mechanisms from a purely cognitive domain to the cognitive-motor domain. This prediction could be empirically tested by asking individuals to imitate a simple reaching movement from various perspectives and measuring their response times.

Therefore, our work suggests that the visuospatial transformation mechanisms embedded in SPL/IPS play a critical role during action observation and imitation by taking into account visuospatial constraints and enabling thus the flexible and robust observation and imitation functionalities of the MNS. Such accuracy, flexibility and robustness in these mechanisms may be particularly critical for fine motor skills such as accurate reaching to grasp a target object while maintaining performance under a variety of conditions. Importantly, it provides a single coherent mechanism that is able to process both the actions of others as well as one's own actions in a common frame of reference. This occurs independently of the imitator-demonstrator visuospatial discrepancies during action observation and imitation, and provides behavioral flexibility and robustness, both of which are critical features of the human motor repertoire.

4.2. Fronto-parietal MNS and its neural dynamics

The bidirectional fronto-parietal MNS circuit (i.e., a loop via pairs of inverse-forward models) is another important feature of our model since it allows a coherent computational process during both observation (action inhibition and use of the efference copy) and imitation (release of neural drives for execution of the actual action). Thus, the fronto-parietal MNS circuits can mimic synaptic patterns that are consistent, both in terms of the network output and synaptic activity, with the IFG and IPL dynamics observed during action observation and imitation. This is in agreement with prior work that demonstrated similar neural activity in the frontal and parietal MNS during action observation and imitation (Arnstein et al., 2011; Gazzola & Keysers, 2009; Grèzes et al., 2003; Turella et al., 2009). The slightly higher IFG activity during imitation relative to observation of actions along with similar activity in the parietal networks (SPL/IPS and IPL) is also consistent with prior work, although such an observation-imitation discrepancy observed in IFG was very small (Nishitani & Hari, 2002). Our results reinforce the previous idea that the bidirectional fronto-parietal networks are engaged not only in mental simulation during self-intended actions where both imagined and actual movements are known to exhibit similar behavioral and neural patterns (Gentili, Oh et al., 2015; Gentili & Papaxanthis, 2015; Gentili et al., 2016; Gueugneau et al., 2015; Guillot, Di Rienzo, Macintyre, Moran, & Collet, 2012; Hétu et al., 2013), but also during action observation/imitation which leads to similar frontal and parietal MNS activity during both observation and imitation of the same action (Jeannerod, 2001, 2006; Vogt & Thomaschke, 2007). Thus our work supports the idea that motor imagery and action observation share, to some extent, common cognitive-motor mechanisms, such as the pairing of inverse-forward computations, action inhibition and efference copy (Gentili, Oh et al., 2015; Gentili et al., 2016; Gueugneau et al., 2015; Gueugneau et al., 2016; Hétu et al., 2013; Jeannerod, 2001; Vogt & Thomaschke, 2007).

Moreover, our model suggests that the visuospatial transformation mechanisms embedded in the SPL/IPS allow the emergence of view-independent neural populations that contribute to computing the neural drive needed to accurately guide the hand to reach and grasp the target object as observed earlier under various viewpoints. Thus, these view-independent neural populations would activate regardless of the perspective from which the action was initially perceived. Such mechanisms could explain findings from prior studies conducted in primates that suggest the existence of view-independent F5 mirror neurons that fire with a similar pattern independently of the demonstrator's position during hand grasping (Caggiano et al., 2011; Caggiano et al., 2015; Craighero, Metta, Sandini, & Fadiga, 2007; Gallese et al., 1996). Also, the presence of view-dependent neural populations in IFG is somewhat consistent with empirical work where neurons with similar features were identified in the F5 area of monkeys (Caggiano et al., 2011; Caggiano et al., 2015). Critically, our model predicts that IPL mainly includes view-independent neurons whereas SPL/IPS would embed both view-independent and view-dependent neurons (likely due to visual motion processing of the observed action of the hand/arm via the MT/MST). This could also be empirically assessed by examining the neural activity in these specific brain regions during action observation/imitation.

Such view-dependent and view-independent neural populations were not preprogrammed into our model, but are an emergent property that results from modeling the computational mechanisms embedded in SPL/IPS and MT/MST. While the IFG, IPL and SPL/IPS were modeled based on empirical evidence and the ratio between view-independent and view-dependent neurons simulated in IFG is similar to that obtained by Caggiano and colleagues during hand grasping (simulated vs empirical VI/VD ratio: 13/87% vs. 26/74%), likely the addition of architectural constraints and brain regions in the model could refine these results (Caggiano et al., 2011). Moreover, although those synaptic patterns can reproduce similar fronto-parietal dynamics in observation/imitation and view-based neural activities can be observed, our MNS model simulates human brain networks' synaptic activity, whereas most past empirical data were obtained from nonhuman primates and while existing human data were obtained via EEG/MEG and fMRI, limiting thus the possibilities for making direct comparisons and/or predictions.

4.3. Comparison to other MNS models

The most important difference between our neural model and other existing MNS models is the modeling of a visuospatial transformation remapping mechanism. In particular, we hypothesize that this transformation system is embedded in the SPL/IPS to support the MNS by providing view-independent representations of observed actions through the fronto-parietal network. As such,

the present work complements previously proposed MNS models, which have not taken into account neural components that learn visuospatial transformations between the demonstrator and the imitator (Bonaiuto & Arbib, 2010; Bonaiuto et al., 2007; Demiris & Hayes, 2002; Demiris & Johnson, 2003; Oztop & Arbib, 2002; Oztop et al., 2005). Another difference is that while our neural model shares with other past MNS modeling efforts an internal model framework, the network structures involved and their functional roles are different. Specifically, Demiris and Hayes (2002), Demiris and Johnson (2003) as well as Tani, Ito, and Sugita (2004) have proposed approaches that are similar to ours in that they combine inverse and forward models. However, while these past studies focused on learning new behaviors to classify observed actions, ours examines the functional role of visuospatial transformation systems as well as MNS activities under various viewpoints during learning by observation and imitation. Furthermore, while a self-organized map was employed to encode the distributed representation of actions in Tani's model (2004), here the RBF network was used to implement a regional representation of actions. However, recent work on self-organizing maps could provide a basis on which the current model could be modified to include self-organizing maps (Huang, Gentili, Katz, & Reggia, 2016; Huang, Gentili, & Reggia, 2015a, 2015b).

Another important distinguishing feature of our model is the original learning strategy which employs online learning while combining both observation and imitation. Such a learning procedure enhances the ecological validity of our approach, since when considering humans: i) the notion that both inverse and forward computations would be learned together rather than sequentially (forward then inverse) as in prior modeling works, is more realistic, and ii) typically the imitator will first observe the action shown by a demonstrator and then will subsequently attempt to reproduce it. Interestingly, our learning scheme resulted in three emergent behaviors that are consistent with earlier studies: i) even an imperfect forward model can train an inverse model (Jordan & Rumelhart, 1992); ii) the forward model tends to be learned earlier than the inverse model, in agreement with the idea that prediction precedes control (Flanagan, Vetter, Johansson, & Wolpert, 2003), and iii) both internal and peripheral feedback are employed to regulate actions (Crevecoeur, Munoz, & Scott, 2016; Crevecoeur & Scott, 2014; Desmurget & Grafton, 2000; Molina-Vilaplana & Coronado, 2006; Molina-Vilaplana et al., 2007). However, our neural model was trained with a supervised learning method, which has limited biological validity compared to unsupervised and reinforcement learning due to the absence of any known physiological mechanisms for rapid non-local backpropagation transport of error signals between synapses (Harris, 2008).

Other interesting MNS models have been developed to emphasize specifically grasp-related mirror neuron activity (Oztop & Arbib, 2002) and to infer the mental states of demonstrators during action observation (Oztop et al., 2005). In their work, Oztop et al. (2005) strongly emphasized the functional role of the PFC in inferring others' mental states (or intentions) when they perform an action, whereas our model mainly uses the rPFC for inhibiting the neural commands and control switching of the visuospatial processes. The model of SPL/IPS-MNS circuits proposed here could complement the mental states inference model in which the intention or goal of action is assumed to be represented in a view-independent manner (Carr et al., 2003; Fogassi et al., 2005).

Until now, only a few computational MNS studies have accounted, to various extents, for viewpoint transformation of observed actions. For instance, Lopes and Santos-Victor (2005) proposed a computational model combining the visuomotor map (i.e., F5 mirror neurons) and viewpoint transformation for learning by imitation. However, their viewpoint transformation model is implemented by a pure transformation matrix without any biological relevance or sensorimotor learning processes. More recently, Arie and colleagues (Arie et al., 2012) proposed a MNS model where novel actions could be successfully imitated by employing a unified intentional representation via the MNS even when observed and self-generated actions differed. However, their findings also revealed that their neural model was able to process angular changes between the imitator and the demonstrator only in a very limited range ($\pm$ 15°). In contrast our model of SPL/IPS regions processes *any* angular variations. Thus, our model complements prior MNS models by incorporating a robust visuospatial transformation system that processes the demonstrator-imitator frame of reference remapping, leading to learning by observation/imitation regardless of visuospatial constraints. In addition, prior modeling work relevant to this study has proposed neurocomputational mechanisms to implement viewpoint transformations (Sauser & Billard, 2005a, 2005b, 2007; Fleischer et al., 2013). Our model complements these prior efforts by examining the relationships between the mechanisms underlying the transformation of the frame of reference and the fronto-parietal MNS, informing thus the interactions between visual processing of actions and motor representations during action observation/imitation. In particular, our model allows for examining the emergence of view-based representations in the frontal-parietal MNS and the parietal regions during learning that may embed viewpoint transformation processes. Although these previous models employ a different approach they are also compatible. For instance, the model proposed by Sauser and Billard that has simulated a 3D viewpoint transformation via a population vector in the STS (Sauser & Billard, 2005a, 2005b, 2007) could be employed to extend our neural model where the STS does not include such a feature.

4.4. Limitations and future work

Although our neural model contributes to understanding the underlying fronto-parietal dynamics during action observation and imitation, it has several limitations that could be addressed in the future. First, the implementations of the PFC and STS could be expanded to allow further exploration of the intention encoding that underlies observed actions and more generally to examine the role of executive functions in adaptive human motor behavior. Second, the synaptic activity computed for each network only allowed a limited assessment of the view-based neurons which could be further examined by considering more advanced computational processing, such as synthetic neuroimaging, permitting a better comparison with human data. Third, due to a prohibitive computational cost, the environment had to be subdivided into smaller local workspaces to train online the three neural networks. However, a single neural network for each of the SPL/IPS, IFG and IPL could learn the entire environment if more computational resources are available in the future. Future work could also expand the model's ability to support 3D actions although it has been suggested that

the recognition of goal-directed actions and visual tuning properties are based on learned view-specific neural representations without the need for an accurate reconstruction of the 3D structure of the effector and thus may be mainly based on 2D views (Wu & Huang, 1999; Erol, Bebis, Nicolescu, Boyle, & Twombly, 2005; Fleischer et al., 2013). Fourth, other neural elements not currently modeled, such as the canonical neurons system that have mirror-like neuron properties could also be included to implement object manipulations during action imitation (Grèzes et al., 2003). In the same vein, although the hypothesized functions of the neural regions considered here were modeled based on strong experimental evidence, various alternative hypotheses regarding their functional role as well as further architectural constraints could be tested in the future.

Also, due to complex dynamics resulting from the simultaneous learning of the SPL/IPS, IFG, and IPL networks, supervised learning was employed to allow us to achieve training more efficiently while mitigating computational complexity. However, such supervised learning techniques (as are used in most of the currently available MNS models) are less biologically valid than unsupervised and reinforcement learning methods since there is no known physiological mechanism for rapid non-local back-propagation transport of error signals between synapses (Harris, 2008). In the same vein, although consideration of having the SPL/IPS regions implemented with four primitive networks to perform various transformations was guided by neurophysiological human evidence, the learning approach, which consisted of learning first these primitive networks followed by the training of SPL/IPS, IFG, and IPL networks, may not be biologically plausible since likely all of these networks are learned altogether. Future work could consider alternative learning approaches (e.g., unsupervised and/or reinforcement learning; learning of the SPL/IPS network primitives directly with IFG and IPL) to further increase the validity of the proposed model (Han, 2009; Han et al., 2008; Huang et al., 2015; Huang et al., 2016). Potential future directions also include additional assessment of our neural model by employing a humanoid robotic system to further test the critical components of this model in a physically realistic environment. Overall, this work aims to contribute to developing a computational platform for studying human adaptive cognitive-motor control (e.g., action observation and imitation, motor imagery, executive functioning) as well as to developing applications relevant to human-robot interactions in a real-world environment.

Acknowledgments

This research was supported in part by the Graduate School Interdisciplinary Dissertation Fellowship from the Graduate School of the University of Maryland, and in part by the Office of Naval Research award N000141310597 and a grant from Lockheed Martin Corporation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.humov.2018.05.013.

References

Abend, W., Bizzi, E., & Morasso, P. (1982). Human arm trajectory formation. *Brain, 105*(Pt 2), 331–348.
Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*(5892), 523–525.
Andersen, R. A. (1987). Inferior parietal lobule function in spatial perception and visuomotor integration. In F. Plum (Ed.). *Handbook of physiology. Section 1: The nervous system. Volume V, Parts 1 & 2: Higher functions of the brain* (pp. 483–518). (1st ed.). Bethesda, MD: American Physiological Society.
Arie, H., Arakaki, T., Sugano, S., & Tani, J. (2012). Imitating others by composition of primitive actions: A neuro-dynamic model. *Robotics and Autonomous Systems, 60*, 729–741.
Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., & Gazzola, V. (2011). Mu-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *Journal of Neuroscience, 31*(40), 14243–14249.
Aziz-Zadeh, L., & Ivry, R. B. (2009). The human mirror neuron system and embodied representations. *Advances in Experimental Medicine and Biology, 629*, 355–376.
Barraclough, N. E., Keith, R. H., Xiao, D., Oram, M. W., & Perrett, D. I. (2009). Visual adaptation to goal-directed hand actions. *Journal of Cognitive Neuroscience, 21*, 1806–1820.
Bisio, A., Avanzino, L., Gueugneau, N., Pozzo, T., Ruggeri, P., & Bove, M. (2015). Observing and perceiving: A combined approach to induce plasticity in human motor cortex. *Clinical Neurophysiology, 126*(6), 1212–1220.
Bisley, J. W., & Pasternak, T. (2000). The multiple roles of visual cortical areas MT/MST in remembering the direction of visual motion. *Cerebral Cortex, 10*(11), 1053–1065.
Blakemore, S. J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research, 153*(2), 239–245.
Bock, O. L., & Dalecki, M. (2015). Mental rotation of letters, body parts and scenes during whole-body tilt: Role of a body-centered versus a gravitational reference frame. *Human Movement Science, 40*, 352–358.
Bonaiuto, J., & Arbib, M. A. (2010). Extending the mirror neuron system model, II: What did I just do? A new role for mirror neurons. *Biological Cybernetics, 102*, 341–359.
Bonaiuto, J., Rosta, E., & Arbib, M. A. (2007). Extending the mirror neuron system model, I. Audible actions and invisible grasps. *Biological Cybernetics, 96*, 9–38.
Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28*, 157–189.
Broomhead, D. S., & Lowe, D. (1988b). Radial basis functions, multi-variable functional interpolation and adaptive networks. London, UK.
Broomhead, D. S., & Lowe, D. (1988a). Multivariable functional interpolation and adaptive networks. *Complex Systems, 2*(3), 321–355.
Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... Freund, H.-J., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13*(2), 400–404.
Buch, E. R., Young, S., & Contreras-Vidal, J. L. (2003). Visuomotor adaptation in normal aging. *Learning & Memory, 10*(1), 55–63.
Bullock, D., Bongers, R. M., Lankhorst, M., & Beek, P. J. (1999). A vector-integration-to-endpoint model for performance of viapoint movements. *Neural Networks, 12*(1), 1–29.
Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review, 95*(1), 49–90.
Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A self-organizing neural model for motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive*

Neuroscience, 5(4), 408–435.
Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia, 44*(13), 2594–2606.
Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences, 11*, 290–298.
Burton, L. A., Wagner, N., Lim, C., & Levy, J. (1992). Visual field differences for clockwise and counterclockwise mental rotation. *Brain and Cognition, 18*, 192–207.
Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Their, P., et al. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Current Biology, 21*, 144–148.
Caggiano, V., Giese, M., Thier, P., & Casile, A. (2015). Encoding of point of view during action observation in the local field potentials of macaque area F5. *European Journal of Neuroscience, 41*(4), 466–476.
Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National academy of Sciences of the United States of America, 100*(9), 5497–5502.
Case, L. K., Pineda, J., & Ramachandran, V. S. (2015). Common coding and dynamic interactions between observed, imagined, and experienced motor and somatosensory activity. *Neuropsychologia, 79*(Pt B), 233–245.
Chen, S., Cowan, C. F. N., & Grant, P. M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks, 2*(2), 302–309.
Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., ... Belliveau, J. W., et al. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain, 119*(1), 89–100.
Craighero, L., Metta, G., Sandini, G., & Fadiga, L. (2007). The mirror-neurons system: Data and models. *Progress in Brain Research, 164*, 39–59.
Crevecoeur, F., Giard, T., Thonnard, J. L., & Lefevre, P. (2011). Adaptive control of grip force to compensate for static and dynamic torques during object manipulation. *Journal of Neurophysiology, 106*, 2973–2981.
Crevecoeur, F., Munoz, D. P., & Scott, S. H. (2016). Dynamic multisensory integration: Somatosensory speed trumps visual accuracy during feedback control. *Journal of Neuroscience, 36*(33), 8598–8611.
Crevecoeur, F., & Scott, S. H. (2014). Beyond muscles stiffness: Importance of state-estimation to account for very fast motor corrections. *PLoS Computational Biology, 10*(10), e1003869.
Crevecoeur, F., Thonnard, J. L., & Lefèvre, P. (2009). Forward models of inertial loads in weightlessness. *Neuroscience, 161*, 589–598.
Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology, 11*, 157–163.
Dalecki, M., Hoffmann, U., & Bock, O. (2012). Mental rotation of letters, body parts and complex scenes: Separate or common mechanisms? *Human Movement Science, 31*(5), 1151–1160.
David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage, 20*(3), 1743–1755.
Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain, 120*(10), 1763–1777.
Demiris, J., & Hayes, G. M. (2002). Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model. In K. Dautenhahn, & C. L. Nehaniv (Vol. Eds.), *Imitation in animals and artifacts*(1st ed.). *vol. 21. Imitation in animals and artifacts* (pp. 327–361). Cambridge, MA: The MIT Press.
Demiris, Y., & Johnson, M. (2003). Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning. *Connection Science, 15*(4), 231–243.
Demiris, Y., & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems, 54*(5), 361–369.
Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron, 37*(2), 347–359.
Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences, 4*(11), 423–431.
Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience, 28*(44), 11231–11239.
Dove, A., Pollmann, S., Schubert, T., Wiggins, C. J., & von Cramon, D. Y. (2000). Prefrontal cortex activation in task switching: An event-related fMRI study. *Cognitive Brain Research, 9*, 103–109.
Duijnhouwer, J., Noest, A., Lankheet, M., Van Den Berg, A., & Van Wezel, R. J. A. (2013). Speed and direction response profiles of neurons in macaque MT and MST show modest constraint line tuning. *Frontiers in Behavioral Neuroscience, 7*, 22.
Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X. (2005) A review on vision-based full DOF hand motion estimation. Paper presented at IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, June.
Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *J Neurophys, 73*(6), 2608–2611.
Flanagan, J. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology, 13*(2), 146–150.
Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience, 33*(15), 6563–6580.
Floreano, D., Ijspeert, A. J., & Schaal, S. (2014). Robotics and neuroscience. *Cell Biology, 24*(18), 910–920.
Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science, 308*(5722), 662–667.
Frank, A. U. (1998). Formal models for cognition – Taxonomy of spatial location description and frames of reference. In C. Freksa, C. Habel, & K. F. Wender (Vol. Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*(1st ed.). *vol. 1404. Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (pp. 293–312). Berlin, Germany: Springer-Verlag Berlin Heidelberg.
Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(Pt 2), 593–609.
Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19*, 1239–1255.
Gentili, R. J., Bradberry, T. J., Oh, H., Costanzo, M. E., Kerick, S. E., Contreras-Vidal, J. L., et al. (2015). Evolution of cerebral cortico-cortical communication during visuomotor adaptation to a cognitive-motor executive challenge. *Biological Psychiatry, 105*, 51–65.
Gentili, R. J., Bradberry, T. J., Oh, H., Hatfield, B. D., & Contreras-Vidal, J. L. (2011). Cerebral cortical dynamics during visuomotor transformation: Adaptation to a cognitive-motor executive challenge. *Psychophysiology, 48*(6), 813–824.
Gentili, R. J., Cahouet, V., & Papaxanthis, C. (2007). Motor planning of arm movements is direction-dependent in the gravity field. *Neuroscience, 145*(1), 20–32.
Gentili, R. J., Oh, H., Huang, D.-W., Katz, G. E., Miller, R. H., & Reggia, J. A. (2015). A neural architecture for performing actual and mentally simulated movements during self-intended and observed bimanual arm reaching movements. *International Journal of Social Robotics, 7*(3), 371–392.
Gentili, R. J., Oh, H., Kregling, V. A., & Reggia, J. A. (2016). A cortical model for inverse kinematics computation of a humanoid finger with mechanically coupled joints. *Bioinspiration & Biomimetics, 11*(3), 036013.
Gentili, R. J., & Papaxanthis, C. (2015). Laterality effects in motor learning by mental practice in right-handers. *Neuroscience, 297*, 231–242.
Gogos, A., Gavrilescu, M., Davison, S., Searle, K., Adams, J., Rossell, S. L., ... Egan, G. F., et al. (2010). Greater superior than inferior parietal lobule activation with increasing rotation angle during mental rotation: An fMRI study. *Neuropsychologia, 48*, 529–535.
Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215.
Gordon, J., Ghilardi, M. F., & Ghez, C. (1994). Accuracy of planar reaching movements. I. Independence of direction and extent variability. *Experimental Brain Research, 99*(1), 97–111.
Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography 2. Observation compared with imagination. *Experimental Brain Research, 112*(1), 103–111.
Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy, 207*(1), 3–17.
Grefkes, C., Ritzl, A., Zilles, K., & Fink, G. R. (2004). Human medial intraparietal cortex subserves visuomotor coordinate transformation. *Neuroimage, 23*(4),

1494–1506.
Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: An fMRI study. *NeuroImage, 18*(4), 928–937.
Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1–19.
Guenther, F. H., & Micci-Barreca, D. (1997). Neural models for flexible control of redundant systems. In P. G. Morasso, & V. Sanguineti (Eds.). *Self-organization, computational maps, and motor control* (pp. 383–421). North Holland: Elsevier.
Gueugneau, N., Bove, M., Ballay, Y., & Papaxanthis, C. (2016). Interhemispheric inhibition is dynamically regulated during action observation. *Cortex, 78*, 138–149.
Gueugneau, N., McCabe, S. I., Villalta, J. I., Grafton, S. T., & Della-Maggiore, V. (2015). Direct mapping rather than motor prediction subserves modulation of corticospinal excitability during observation of actions in real time. *Journal of Neurophysiology, 113*(10), 3700–3707.
Gueugneau, N., Schweighofer, N., & Papaxanthis, C. (2015). Daily update of motor predictions by physical activity. *Scientific Reports, 3*(5), 17933.
Guillot, A., Di Rienzo, F., Macintyre, T., Moran, A., & Collet, C. (2012). Imagining is not doing but involves specific motor commands: A review of experimental data related to motor inhibition. *Frontiers in Human Neuroscience, 6*(247).
Han, C. (2009). *Modeling human reaching and grasping: Cortex, rehabilitation and lateralization.* University of Southern California, ProQuest Dissertations Publishing.
Han, C. E., Arbib, M. A., & Schweighofer, N. (2008). Stroke rehabilitation reaches a threshold. *PLoS Computational Biology, 4*(8), e1000133.
Harris, K. D. (2008). Stability of the fittest: Organizing learning through retroaxonal signals. *Trends in Neuroscience, 31*(3), 130–136.
Herdin, M., Czink, N., Ozcelik, H., & Bonek, E. (2005). Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. *2005 IEEE 61st vehicular technology conference. vol. 1. 2005 IEEE 61st vehicular technology conference* (pp. 136–140). Piscataway, NJ: IEEE.
Hesse, M. D., Sparing, R., & Fink, G. R. (2009). End or means – The "what" and "how" of observed intentional actions. *Journal of Cognitive Neuroscience, 21*(4), 776–790.
Hétu, S., Grégoire, M., Saimpont, A., Coll, M. P., Eugène, F., Michon, P. E., et al. (2013). The neural network of motor imagery: An ALE meta-analysis. *Neuroscience & Biobehavioral Reviews, 37*, 930–949.
Huang, D.-W., Gentili, R.J., & Reggia, J.A. (2015a). A self-organizing map architecture for arm reaching based on limit cycle attractors. In 9th EAI international conference on bio-inspired information and communications technologies (BICT'15), December 3–5, p. 1–6, New York City, NY, USA.
Huang, D. W., Gentili, R. J., Katz, G. E., & Reggia, J. A. (2016). A limit-cycle self-organizing map architecture for stable arm control. *Neural Networks, 85*, 165–181.
Huang, D. W., Gentili, R. J., & Reggia, J. A. (2015b). Self-organizing maps based on limit cycle attractors. *Neural Networks, 63*, 208–222.
Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology, 15*(6), 632–637.
Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., ... Rizzolatti, G., et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 98*(24), 13995–13999.
Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526–2528.
Ilg, U. J. (2008). The role of areas MT and MST in coding of visual motion underlying the execution of smooth pursuit Author links open overlay panel. *Vision Research, 48*(20), 2062–2069.
Ito, M. (2002). Historical review of the significance of the cerebellum and the role of Purkinje cells in motor learning. *Annals of the New York Academy of Sciences, 978*, 273–288.
Jansen, B. H., & Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics, 73*(4), 357–366.
Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage, 14*(1 Pt 2), 103–109.
Jeannerod, M. (2006). The origin of voluntary action. History of a physiological concept. *Comptes Rendus Biologies, 329*, 354–362.
Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16*, 307–354.
Jordan, M., & Wolpert, D. M. (1999). Computational motor control. In Gazzaniga (Ed.). *The Cognitive Neurosciences*(pp. 601–620). Cambridge, MA: MIT Press.
Kadmon Harpaz, N., Flash, T., & Dinstein, I. (2014). Scale-invariant movement encoding in the human motor system. *Neuron, 81*(2), 452–461.
Kagerer, F. A., Contreras-Vidal, J. L., & Stelmach, G. E. (1997). Adaptation to gradual as compared with sudden visuo-motor distortions. *Experimental Brain Research, 115*(3), 557–561.
Katz, G. E., Huang, D.-W., Hauge, T. C., Gentili, R. J., & Reggia, J. A. (2017). *A novel parsimonious cause-effect reasoning algorithm for robot imitation and plan recognition.* IEEE Transactions on Cognitive and Developmental Systems: In press.
Kilner, J., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*, 159–166.
Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506–1509.
Li, C. S. R., Padoa-Schioppa, C., & Bizzi, E. (2001). Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron, 30*, 593–607.
Lopes, M., & Santos-Victor, J. (2005). Visual learning by imitation with motor representations. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, 35*(3), 438–449.
Meyer, D.E., Evans, J.E., Lauber, E.J., Rubinstein, J., Gmeindl, L., Junck, L., & Koeppe, R.A. (1997). *Activation of brain mechanisms for executive mental processes in cognitive task switching.* Ann Arbor, MI.
Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport, 14*(17), 2135–2137.
Miller, A. J. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society. Series A (General), 147*(3), 389–425.
Miller, E., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.
Molina-Vilaplana, J. M., & Coronado, J. L. (2006). A neural network model for coordination of hand gesture during reach to grasp. *Neural Networks, 19*(1), 12–30.
Molina-Vilaplana, J. M., Feliu-Batlle, J., & López-Coronado, J. (2007). A modular neural network architecture for step-wise learning of grasping tasks. *Neural Networks, 20*(5), 631–645.
Morasso, P. (1981). Spatial control of arm movements. *Experimental Brain Research, 42*(2), 223–227.
Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Proceedings of the National Academy of Sciences of the United States of America, 97*(2), 913–918.
Nishitani, N., & Hari, R. (2002). Viewing lip forms: Cortical dynamics. *Neuron, 36*(6), 1211–1220.
Oh, H., Gentili, R.J., Reggia, J.A., & Contreras-Vidal, J.L. (2011). Learning of spatial relationships between observed and imitated actions allows invariant inverse computation in the frontal mirror neuron system. In *Proceedings of annual international conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4183–4186.
Oh, H., Gentili, R.J., Reggia, J.A., & Contreras-Vidal, J.L. (2012). Modeling of visuospatial perspectives processing and modulation of the fronto-parietal network activity during action imitation. In *Proceedings of annual international conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2551–2554.
Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: An MVPA Study. *Journal of Cognitive Neuroscience, 24*(4), 975–989.
Orr, M. (1998). Optimising the widths of radial basis functions. In A. de P. Braga, & T. B. Ludermir (Eds.). *Proceedings 5th Brazilian symposium on neural networks* (pp. 26–29). Belo Horizonte, Brazil: IEEE Computer Society.
Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics, 87*(2), 116–140.
Oztop, E., Bradley, N. S., & Arbib, M. A. (2004). Infant grasp learning: A computational model. *Experimental Brain Research, 158*, 480–503.
Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks, 19*(3), 254–271.
Oztop, E., Kawato, M., & Arbib, M. A. (2013). Mirror neurons: Functions, mechanisms and models. *Neuroscience Letters, 540*, 43–55.
Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research, 22*(2), 129–151.
Padoa-Schioppa, C., Li, C. S., & Bizzi, E. (2004). Neuronal activity in the supplementary motor area of monkeys adapting to a new dynamic environment. *Journal of Neurophysiology, 91*(1), 449–473.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience, 21*, 227–277.
Peuskens, H., Sunaert, S., Dupont, P., Van Hecke, P., & Orban, G. A. (2001). Human brain regions involved in heading estimation. *Journal of Neuroscience, 21*(7), 2451–2461.
Popa, L. S., Hewitt, A. L., & Ebner, T. J. (2012). Predictive and feedback performance errors are signaled in the simple spike discharge of individual purkinje cells. *Journal of Neuroscience, 32*(44), 15345–15358.
Porrill, J., & Dean, P. (2007). Recurrent cerebellar loops simplify adaptive control of redundant and nonlinear motor systems. *Neural Computation, 19*(1), 170–193.
Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology, 210*, 1–3.
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.
Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience, 2*(9), 661–670.
Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology, 12*, 149–154.
Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: New concepts. *Electroencephalography and Clinical Neurophysiology, 106*(4), 283–296.
Rogers, R. D., Sahakian, B. J., Hodges, J. R., Polkey, C. E., Kennard, C., & Robbins, T. W. (1998). Dissociating executive mechanisms of task control following frontal lobe damage and Parkinson's disease. *Brain, 121*(5), 815–842.
Sauser, E. L., & Billard, A. G. (2005a). Three-dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing, 64*, 5–24.
Sauser, E. L., & Billard, A. G. (2005b). View sensitive cells as a neural basis for the representation of others in a self-centered frame of reference. *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts* (pp. 119–127). UK: Hatfield.
Sauser, E., & Billard, A. (2007). Interferences in the transformation of reference frames during a posture imitation task. *Proceeding of the International Conference on Artificial Neural Networks*. Porto, Portugal: ICANN 07.
Schaal, S., & Schweighofer, N. (2005). Computational motor control in humans and robots. *Current Opinion in Neurobiology, 15*(6), 675–682.
Schwabe, L., Lenggenhager, B., & Blanke, O. (2009). The timing of temporoparietal and frontal activations during mental own body transformations from different visuospatial perspectives. *Human Brain Mapping, 30*(6), 1801–1812.
Schweighofer, N., Lang, E. J., & Kawato, M. (2013). Role of the olivo-cerebellar complex in motor learning and control. *Frontiers in Neural Circuits, 28*(7), 94.
Shmuelof, L., & Zohary, E. (2008). Mirror-image representation of action in the anterior parietal cortex. *Nature Neuroscience, 11*(11), 1267–1269.
Sirigu, A., Duhamel, J.-R., Cohen, L., Pillon, B., Dubois, B., & Agid, Y. (1996). The mental representation of hand movements after parietal cortex damage. *Science, 273*(5281), 1564–1568.
Srinivasa, N., Bhattacharyya, R., Sundareswara, R., Lee, C., & Grossberg, S. (2012). A bio-inspired kinematic controller for obstacle avoidance during reaching tasks with real robots. *Neural Networks, 35*, 54–69.
Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62*(3), 626–641.
Tani, J., Ito, M., & Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot experiments using RNNPB. *Neural Networks, 17*(8–9), 1273–1289.
Tin, C., & Poon, C.-S. (2005). Internal models in sensorimotor integration: Perspectives from adaptive control theory. *Journal of Neural Engineering, 2*, 147–163.
Tokuda, I. T., Han, C. E., Aihara, K., Kawato, M., & Schweighofer, N. (2010). The role of chaotic resonance in cerebellar learning. *Neural Networks, 23*(7), 836–842.
Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience, 15*(4), 3215–3230.
Turella, L., Erb, M., Grodd, W., & Castiello, U. (2009). Visual features of an observed agent do not modulate human brain activity during action observation. *Neuroimage, 46*, 844–853.
Vogt, S., & Thomaschke, R. (2007). From visuomotor interactions to imitation learning: Behavioural and brain imaging studies. *Journal of Sports Sciences, 25*, 497–517.
Wolpert, D. M., Goodbody, S. J., & Husain, M. (1998). Maintaining internal representations: The role of the human superior parietal lobe. *Nature Neuroscience, 1*(6), 529–533.
Wolpert, D. M., & Miall, R. C. (1996). Forward models for physiological motor control. *Neural Networks, 9*, 1265–1279.
Wu, Y., & Huang, T. S. (1999). Vision-based gesture recognition: a review. In A. Braffort, R. Gherbi, S. Gibet, J. Richardson, & E. Teil (Eds.). *Proceedings of the international gesture workshop on gesture-based communication in human-computer interaction* (pp. 103–115). London: Springer-Verlag.
Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19.