ABSTRACT

Title of Document:	A MULTIPLE REPRESENTATIONS MODEL OF THE HUMAN MIRROR NEURON SYSTEM FOR LEARNED ACTION IMITATION.
	Hyuk Oh, Doctor of Philosophy, 2015
Directed By:	Assistant Professor, Rodolphe J. Gentili, Neuroscience and Cognitive Science Program

The human mirror neuron system (MNS) is a fundamental sensorimotor system that plays a critical role in action observation and imitation. Despite a large body of experimental and theoretical MNS studies, the visuospatial transformation between the observed and the imitated actions has received very limited attention. Therefore, this work proposes a neurobiologically plausible MNS model, which examines the dynamics between the fronto-parietal mirror system and the parietal visuospatial transformation system during action observation and imitation. The fronto-parietal network is composed of the inferior frontal gyrus (IFG) and the inferior parietal lobule (IPL), which are postulated to generate the neural commands and the predictions for its sensorimotor consequences, respectively. The parietal regions identified as the superior parietal lobule (SPL) and the intraparietal sulcus (IPS) are postulated to encode the visuospatial transformation for enabling view-independent representations of the observed action. The middle temporal region is postulated to provide the view-dependent representations such as direction and velocity

of the observed action. In this study, the SPL/IPS, IFG, and IPL are modeled with artificial neural networks to simulate the neural mechanisms underlying action imitation. The results reveal that this neural model can replicate relevant behavioral and neurophysiological findings obtained from previous action imitation studies. Specifically, the imitator can replicate the observed actions independently of the spatial relationships with the demonstrator while generating similar synthetic functional magnetic resonance imaging blood oxygenation level-dependent responses in the IFG for both action observation and execution. Moreover, the SPL/IPS can provide view-independent visual representations through mental transformation for which the response time monotonically increases as the rotation angle augments. Furthermore, the simulated neural activities reveal the emergence of both view-independent and view-dependent neural populations in the IFG. As a whole, this work suggests computational mechanisms by which visuospatial transformation processes would subserve the MNS for action observation and imitation independently of the differences in anthropometry, distance, and viewpoint between the demonstrator and the imitator.

A MULTIPLE REPRESENTATIONS MODEL OF THE HUMAN MIRROR NEURON SYSTEM FOR LEARNED ACTION IMITATION.

By

Hyuk Oh.

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Assistant Professor Rodolphe J. Gentili, Chair
Acting Chief Allen R. Braun
Professor Bradley D. Hatfield
Professor James A. Reggia
Professor William J. Idsardi, Dean's Representative

Acknowledgements

It is a great honor for me to work with my academic advisor and doctoral committees. I have received a great support and encouragement from them. I would like to thank my academic advisor, Dr. Rodolphe Gentili for his great mentorship and friendship. I could not complete my doctorate without invaluable insights from all my dissertation committees of Drs. Allen Braun, Bradley Hatfield, William Idsardi, and James Reggia. Finally, I would like to give special thanks to my family for their continuous and unconditional love, encouragement, and support.

This research was supported in part by the Graduate School Interdisciplinary Dissertation Fellowship from the Graduate School of the University of Maryland (UMD) and in part by the Office of Naval Research (ONR) award N000141310597. The content is solely the responsibility of the author and does not necessarily represent the official views of the UMD or ONR.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbr.	Definition	Abbr.	Definition	Abbr.	Definition
ASD	Autism Spectrum Disorders	**BOLD**	Blood Oxygenation Level-Dependent	**CB**	Cerebellum
(r)CBF	(regional) Cerebral Blood Flow	**CBV**	Cerebral Blood Volume	**CMD**	Covariance Matrix Distance
CMRO2	Cerebral Metabolic Rate of Oxygen	**CNS**	Central Nervous System	**CSF**	Cerebrospinal Fluid
CCW	Counterclockwise	**CW**	Clockwise	**EEG**	Electroencephalography
FIR	Finite Impulse Response	**GCV**	Generalized Cross-Validation	**GLM**	General Linear Model
Hb	Hemoglobin	**HbO2**	Oxyhemoglobin	**HbR**	Deoxyhemoglobin
HR(F)	Hemodynamic Response (Function)	**IFG**	Inferior Frontal Gyrus	**IMA**	Ideomotor Apraxia
IPL	Inferior Parietal Lobule	**IPS**	Intraparietal Sulcus	**ISI**	Interstimulus Interval
ITI	Intertrial Interval	**LFP**	Local Field Potential	**LTI**	Linear Time-Invariant System
MEG	Magnetoencephalography	**MNS**	Mirror Neuron System	**MOSAIC**	Modular Selection And Identification for Control Model
MR	Magnetic Resonance	**(f)MRI**	(functional) Magnetic Resonance Imaging	**MSE**	Mean Squared Error
MSI	Mental State Inference Model	**MT**	Middle Temporal	**OLS**	Orthogonal Least Square
PET	Positron Emission Tomography	**(r)PFC**	(rostral) part of the Prefrontal Cortex	**PPC**	Posterior Parietal Cortex
RBF	Radial Basis Function	**RF**	Radiofrequency	**RMSE**	Root-Mean-Square Error
SNR	Signal-to-Noise Ratio	**SOA**	Stimulus Onset Asynchrony	**SPL**	Superior Parietal Lobule
SPM	Statistical Parametric Mapping Software	**SSE**	Sum of Squared Errors	**STS**	Superior Temporal Sulcus
TE	Echo Time	**TMS**	Transcranial Magnetic Stimulation	**TR**	Repetition Time

Chapter 1: Introduction

The human imitation faculty has been generally understood to be a primary means of learning complex skills such as facial expressions, manual actions, and language, particularly through imitation studies in infancy (Kuhl & Meltzoff, 1996; Meltzoff & Moore, 1983, 1977). Moreover, cross-species comparisons between humans and nonhuman primates have revealed that humans employ some unique cognitive mechanisms such as counterfactual reasoning in imitation learning, which allows for learning from others' mistakes (Want & Harris, 2001). Such distinctive human ability to reproduce the behaviors observed at an earlier time is called deferred imitation (or observational learning), and plays a key role in passing learned knowledge onto others and even next generations through so-called the diffusion chains (Flynn & Whiten, 2008; McDonough, Mandler, McKee, & Squire, 1995). Therefore, it appears that imitation through observational learning is a highly complex cognitive process employing various cognitive abilities such as visual perception, memory, recall, reproduction, and motivation of behaviors (Carroll & Bandura, 1987). Consequently, the study of imitation learning has become increasingly popular as it could offer a new route to develop our understanding of functional relationships between representations of perception and action, efficient motor learning, and modular motor control.

In 1908, Liepmann initially postulated that humans have particular imitation systems in the left parietal region by using the deficit-lesion method, which corresponds to what is now called ideomotor apraxia (IMA), where the patients with IMA have difficulty with

imitation and performance of meaningful gestures on verbal command (Goldenberg, 2003). In accordance with this pioneering work, subsequent researches have provided supporting evidences that impaired imitation observed in IMA correlates with not only the left parietal cortex but also the left premotor cortex, the supplementary motor area, and the posterior parietal cortex (PPC) including the superior parietal lobule (SPL) and the left inferior parietal lobule (IPL) (Goldenberg, 2009; Imazu, Sugio, Tanaka, & Inui, 2007; Wheaton, Nolte, Bohlhalter, Fridman, & Hallett, 2005). Later, due to the advent of various neurophysiological and brain-imaging technologies, numerous studies have indicated the existence of a large temporo-parieto-frontal network, called the mirror system or mirror neuron system (MNS) (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Giacomo Rizzolatti, Fogassi, & Gallese, 2001). Specifically, many neuroimaging studies have revealed that the inferior frontal gyrus (IFG) and the IPL exhibit greater activation in the context of MNS, thus these two brain regions are named the frontal and parietal MNS, respectively (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Giacomo Rizzolatti et al., 2001). In addition, a so-called mirror-like system in the superior temporal sulcus (STS) has been often considered in MNS studies, because it provides a visual description of the observed action to the parietal MNS (Iacoboni et al., 2001). Specifically, the STS responds to the visual representation of body limbs involved in the observed action (Iacoboni, 2005), but is not activated when unknown actions are perceived (R. Christopher Miall, 2003). In particular, a similar MNS activity during action observation and execution clearly suggests that this network is responsible for

action understanding and imitation learning, since imitation requires the abilities to observe an action, and subsequently, to replicate the observed action (Carr et al., 2003).

Interestingly, when considering action imitation in an ecologically valid context, naïve spectators can easily notice that humans can observe and reproduce other individual's actions independently of various spatial relationships between them (e.g., the differences in distance and viewpoint). Remarkably, several sensorimotor studies have investigated the neural processes underlying mental transformation of visuospatial information such as re-orientation, rotation, and scaling (Buneo & Andersen, 2006; Culham & Kanwisher, 2001; Grefkes & Fink, 2005; Thiel, Zilles, & Fink, 2004; Wolpert, Goodbody, & Husain, 1998). These studies have provided converging evidences that the PPC, particularly the SPL and intraparietal sulcus (IPS), plays a critical role in visuospatial processing. Interestingly, some MNS studies have recently focused on the importance of such visuospatial processes in imitation through observational learning. These studies have provided indirect as well as direct evidences that the MNS functionalities are mediated by view-independent and view-dependent representations of the observed actions (Caggiano et al., 2011; Hesse, Sparing, & Fink, 2009; Oosterhof, Tipper, & Downing, 2012).

Besides such a large body of experimental studies, several conceptual and computational modeling approaches have been proposed to understand the neural mechanisms and functional roles of the MNS, in particular the IFG and IPL that are two key components of it (J. B. Bonaiuto, Rosta, & Arbib, 2007; Iacoboni, 2005; R. Christopher Miall, 2003; Oztop, Wolpert, & Kawato, 2005). Specifically, these models commonly emphasized the sensorimotor control aspects of the MNS in imitation learning, thus they have adopted the theoretical internal model framework to guide such a functional role of the MNS. Within

this internal model framework, it has been presumed that the STS-IPL-IFG pathway would work as an inverse model by creating the motor representation available for imitation from the visual representation of an observed action (Iacoboni et al., 1999). Moreover, it has been suggested that the reverse IFG-IPL-STS pathway would serve as a forward model by building the specified visual representation for a self-action from the corresponding motor representation to be imitated (Iacoboni et al., 1999). Subsequently, this theoretical model has been extended by incorporating the cerebellum (CB) to implement another pair of inverse and forward models in parallel with the existing temporo-parieto-frontal network (R. Christopher Miall, 2003). Interestingly, in the Miall's model, the PPC works as a hub of the model interfacing between the IFG, IPL, and STS. Specifically, the PPC interacts with the MNS network to provide trajectory-invariant visuospatial representations of perception and action (R. Christopher Miall, 2003).

Although these modeling approaches are highly informative to examine mechanisms of the MNS in imitation learning, several important computational elements have received little attention. First, a small minority of studies have focused on adaptive inverse control (i.e., visual-to-motor mapping) that is relevant to the frontal MNS. Second, the parieto-frontal interaction (i.e., forward-inverse coupling between IPL and IFG) has not been taken into consideration in the computational models during imitation through observational learning. In other words, the two-way process of both visual-to-motor and motor-to-visual transformations, which underpins the imitative learning, has not been computationally demonstrated. Third, no computational efforts have attempted to integrate the PPC as a provider of view-invariant representation into the MNS model.

Specifically, for action imitation, the PPC allows the MNS to decode the intentions inherent in the perceived actions by processing any difference in anthropometry (e.g., upper arm and forearm lengths), distance (e.g., close or far), the functional ranges of motion (e.g., shoulder horizontal adduction and elbow horizontal flexion), and more importantly viewpoint (e.g., facing each other or in the same direction) between an imitator and a demonstrator. Fourth, very few studies have validated their neural network models by means of synthetic neuroimaging methods in addition to network and behavioral performances that are typically used in most of the existing computational MNS models and other sensorimotor models (Michael A. Arbib, Billard, Iacoboni, & Oztop, 2000). Namely, a majority of the computational models cannot bridge the growing brain-imaging data obtained in many MNS studies with their proposed model mechanisms. Therefore, in this work, a novel neural architecture is proposed to address these four knowledge gaps one by one with the following four specific aims.

1.1 Specific aims (SA)

1.1.1 Adaptive inverse scheme in IFG

This first SA proposes to model how the brain learns to control the upper limb through inverse computation in the IFG (i.e., frontal MNS) from observed actions. It is predicted that this model can learn to imitate by acquiring the relationship between observed actions and the resulting motor plans used to reproduce such observed actions.

1.1.2 Fronto-parietal interaction between IFG and IPL throughout learning

The second SA addresses how the brain can learn to predict actions through forward computation in the IPL (i.e., parietal MNS), and its interaction with the IFG (i.e., frontal MNS). Such a predictive mechanism is likely to be related to the developmental aspects in the imitation through observational learning. Specifically, the predictions provided by the IPL can be imperfect in the early developmental stage, but they would be still useful to gradually train the inverse computation in the IFG. It is predicted that this coupling between IFG and IPL allows observational learning and reproduction of the observed actions by modeling a chain of mirroring function, which is known to have functionally and computationally similar properties during action observation and execution of the same action. It is also predicted that the formation of the IPL (i.e., forward model) should precedes that of the IFG (i.e., inverse model) throughout the learning.

1.1.3 Visuospatial transformation in SPL and IPS

The third SA examines the potential role of the SPL and IPS in visuospatial transformations from the demonstrator's allocentric to the imitator's egocentric space

required for observational learning and imitation. It is predicted that such a transformation capability allows the imitator to observe the demonstrator's actions without constraints of the frames of reference, viewpoints, and anthropometric data, and in turn, to observe even its own actions for visual feedback. Particularly, this visual feedback is served to improve the quality of the learned actions in both IFG and IPL. It is also expected that the SPL/IPS learns faster than the IFG allowing thus the IFG to correctly perform the visual-to-motor mapping.

1.1.4 Model validation with behavioral and synthetic functional neuroimaging data

Although the mechanisms and functional roles of each MNS component are carefully considered in the development of the MNS model, there still exists the knowledge gap between the simulated neural activity from the MNS model and actual functional neuroimaging data from the experimental findings. Therefore, to bridge this gap, a synthetic functional neuroimaging approach is proposed to generate realistic neural activity patterns that simulate functional magnetic resonance imaging (fMRI) data during action observation and execution. Then, the cross-validation method between the synthetic and actual data under the same condition is employed to assess the validity of the proposed MNS model. It is predicted that after learning of the three networks, similarly to experimental MNS studies, the activity patterns of the MNS will be similar between the observation and execution of the action. Moreover, it is expected that both view-independent and view-dependent neural populations should emerge in the IFG after learning.

Therefore, by proposing a novel neurobiologically plausible MNS architecture, the overall goal of this work is to examine the hypotheses that i) the specific parietal regions (i.e., SPL/IPS) would be critical to implement visuospatial transformation capability, and ii) this transformation system would subserve the MNS (i.e., IFG and IPL) for action observation and imitation independently of the differences in anthropometry, distance, and viewpoint. This proposed MNS model is validated by employing both behavioral and neuroimaging data on the mirror neurons and MNS literatures.

The remainder of this manuscript is organized as follows. The second chapter reviews some representative experimental, theoretical, and computational literatures related to both MNS and synthetic neuroimaging techniques, which guide the developments of the proposed MNS and the synthetic neuroimaging models, respectively. The third chapter presents the methods, which include i) the mechanisms of the proposed conceptual MNS model, ii) the network architecture and learning algorithm to implement this MNS model, iii) the task conditions to train the model, and finally iv) the computational methods to model the corresponding synthetic neuroimaging data to validate the MNS activity. The fourth chapter assesses the performance of the proposed MNS model by investigating various measurements such as kinematics, response time, and neural population activity. Finally, the last chapter provides a summary and a discussion as well as the implications of the current results for future directions of this work.

Chapter 2: Backgrounds and Literature reviews

2.1 Neurophysiological evidence of mirror neurons and mirror systems

2.1.1 Mirror neurons in monkeys

Mirror neurons are specialized neurons originally discovered in macaque monkeys that have the functional properties of discharging when the monkey not only observes specific goal-directed actions performed by another individual, but also executes the same (i.e., strictly congruent neurons) or similar (i.e., broadly congruent neurons) actions (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). These neurons were originally found in the ventral premotor cortex (area F5) of a macaque monkey (Gallese et al., 1996; Giacomo Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), and subsequently also discovered in the rostral part of the IPL (area PF or Brodmann area 7b) (Giacomo Rizzolatti et al., 2001). Although both F5 and PF mirror neurons have aforementioned mirror properties, interestingly, they also have their own intrinsic properties. In particular, F5 mirror neurons discharge when the monkey sees sufficient partial traces of the occurring action to mentally simulate it (Umiltà et al., 2001). Moreover, most of PF mirror neurons selectively fire only when a specific action (e.g., grasping) is followed by a subsequent specific action (e.g., eating or placing in the context of sequential actions with different goals such as grasping for eating or grasping for placing) (Fogassi et al., 2005). These findings indicate that F5 mirror neurons are particularly involved in the

understanding of action, and are postulated to code for motor schemas[1] of actions manipulating objects (Giacomo Rizzolatti et al., 1996, 2001; Umiltà et al., 2001). On the other hand, PF mirror neurons code for the prediction of the next action that is not observed yet as well as for kinesthetic components of actions (Chaminade, Meltzoff, & Decety, 2005), thus they are presumed to be involved in the understanding of intentions inherent in the perceived actions (Carr et al., 2003; Fogassi et al., 2005).

Further experiments have demonstrated that F5 mirror neurons selectively discharge when the monkey observes a biological end-effector (e.g., hand or mouth) interacting with an object (e.g., a reaching and grasping action), but do not discharge at all in response to the sight of only one of them (e.g., a hand action without an object or the simple presentation of an object) (Gallese et al., 1996; Giacomo Rizzolatti et al., 1996). Moreover, it has been found that F5 mirror neurons respond when the monkey not only performs a visually guided specific hand action (e.g., peanut breaking, paper ripping, etc.), but also perceives the corresponding action specific sounds (Kohler et al., 2002). Furthermore, recent studies of both ingestive mouth actions (e.g., such as sucking or breaking food) and facial communicative actions (e.g., tongue and lip protrusion in infant macaque monkeys) have showed that the observation and execution of mouth gestures lead to similar responses in the F5 mirror neurons (Pier F. Ferrari et al., 2006; Pier Francesco Ferrari, Gallese, Rizzolatti, & Fogassi, 2003).

[1] A motor schema is an abstract set of rules for determining a movement that is produced by varying the parameters, which determine a specific movement (Schmidt, 1975). For instance, people produce a movement by manipulating muscle activation with the parameters such as the duration, level, or overall time of force.

Interestingly, it was found that the responses of F5 mirror neurons are unrelated with the identity of the demonstrator, who is a human or a monkey, and are not affected by the body size of the demonstrator (Gallese et al., 1996). More importantly, in the same study, they showed that F5 mirror neurons congruently response to the grasping action regardless of whether this action is performed in the center, on the right side, or on the left side of the monkey. Moreover, a recent study revealed that the F5 mirror neurons could be classified into two classes, view-dependent and view-independent, with respect to the viewpoint from which the actions performed by a demonstrator are observed (Caggiano et al., 2011). Specifically, the majority of the F5 mirror neurons (74%) selectively responds to the viewpoint (i.e., view-dependent mirror neurons), whereas a minority of them (26%) shows response invariance with respect to viewpoints (i.e., view-independent mirror neurons). The existence of view-independent mirror neurons could strengthen the core for a functional role of mirror neurons in action understanding, because it dissociates the higher order visuospatial cognitive processes associated with mental rotation from the possible functions of the mirror neurons (Oh, Gentili, Reggia, & Contreras-Vidal, 2012). In other words, such view-independent mirror neurons match visuomotor representations of the observed actions in terms of their goals independently of the detailed visual characteristics such as a viewpoint. Although the number of view-independent mirror neurons is only about one fourth of the whole F5 mirror neurons (Caggiano et al., 2011), the existence of such neurons is also consistent with a finding in Gallese et al. (1996).

2.1.2 Mirror systems in humans

In humans, direct neurophysiological evidence of mirror properties was originally obtained in the left premotor areas by employing transcranial magnetic stimulation (TMS) (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). Subsequent TMS studies have revealed that the human MNS holds two unique mirror properties not observed in monkeys (Maeda, Kleiner-Fisman, & Pascual-Leone, 2002; Patuzzo, Fiaschi, & Manganotti, 2003). One is that meaningless hand gestures as well as goal-directed hand actions lead to the activation in the human mirror systems, although the former does not activate mirror neurons in monkeys. The other is that the human mirror systems respond to both an action (e.g., a reaching and grasping action) and the movements forming it (e.g., arm reaching, finger pre-shaping, and finger closing movements), whereas the monkey mirror neurons are fired only when observing the whole (i.e., a goal-directed) action (Giacomo Rizzolatti & Craighero, 2004). These findings suggest why humans have higher imitation faculty than other nonhuman primates (Gangitano, Mottaghy, & Pascual-Leone, 2001; Giacomo Rizzolatti & Craighero, 2004).

Since this first description of substantial evidence for human mirror systems, a large number of studies has attempted to identify the human homologue of monkey mirror neurons using various different techniques such as positron emission tomography (PET) (e.g., Grafton, Arbib, Fadiga, & Rizzolatti, 1996), fMRI (e.g., G. Buccino et al., 2001), electroencephalography (EEG) (e.g., Cochin, Barthelemy, Lejeune, Roux, & Martineau, 1998), and magnetoencephalography (MEG) (e.g., Hari et al., 1998). These studies have mainly addressed functional properties of the IFG and IPL, which structurally correspond to the F5 and PF mirror neurons in macaque monkeys, during the execution and

observation of goal-directed actions. Interestingly, their results have commonly supported that each of these two brain areas forms a much wider network than monkey mirror neurons, and that their mirror properties are similar to their macaque monkey counterparts. Moreover, it has been found that observation of egocentric actions (i.e., self-actions) results in a contralateral activation of the IFG and IPL, whereas observation of allocentric actions (i.e., others' actions) generates greater ipsilateral IFG and IPL activation (Aziz-Zadeh, Maeda, Zaidel, Mazziotta, & Iacoboni, 2002; Shmuelof & Zohary, 2007, 2008). On the other hand, during imitation, it has been found that the activity in the IFG and IPL is fairly bilateral, but is stronger in the ipsilateral hemisphere (Aziz-Zadeh, Koski, Zaidel, Mazziotta, & Iacoboni, 2006). Therefore, it seems reasonable to predict that these two brain regions regardless of functional lateralization constitute two cores of the MNS, in which each of them is respectively named the frontal MNS and the parietal MNS (Decety et al., 1997; Fadiga et al., 1995). Besides these two main components, a third component named mirror-like system has been identified in the STS (Iacoboni et al., 2001), which responds to the biological motion of body parts (e.g., face and hands), but does not activate during execution of the unseen action (R. Christopher Miall, 2003). Although the STS does not have motor properties, it plays a crucial role in the action imitation, which are implemented through a temporo-parieto-frontal network (Giacomo Rizzolatti et al., 2001). Another human mirror-like system named the canonical neuron system has been identified in the ventral premotor cortex, which has comparable properties to canonical neurons as its macaque monkey counterpart (Chao & Martin, 2000). In particular, the canonical neuron system is often studied in relationship to the MNS, since it is activated when an individual either

executes or observes an action in the presence of objects that can be manipulated, but does not respond to the observation of objects alone (Chao & Martin, 2000; R. Christopher Miall, 2003). However, the canonical neuron system is not a part of the MNS because it is not activated without an object to be manipulated.

2.1.2.1 Other functional roles of the human mirror systems

As stated in the previous section, the human MNS is primarily responsible for action observation as well as action imitation. In addition to such functional roles, a large number of studies has investigated other important functional roles of the MNS.

2.1.2.1.1 Imitation through observational learning

An event-related fMRI study investigated the MNS activity in three conditions; i) the observation of an action with an explicit instruction to imitate, ii) the observation of an action without an explicit instruction to imitate, and iii) the observation of an action with an explicit instruction not to imitate (Giovanni Buccino et al., 2004). The results showed that the MNS activates in all three conditions, but the strength of the activation is stronger in the first condition (i.e., the observation of an action with an explicit instruction to imitate) than the other two conditions. Interestingly, the SPL is also activated in this first condition, but not in the third condition (i.e., the observation of an action with an explicit instruction not to imitate). This suggests that the MNS is involved in the acquisition of new action sequences through observational learning, and, with the intention to imitate, the SPL interacts closely with the MNS. This result also implies that, in the context of the imitation and intention, the SPL may belong to the controlled cognitive processes rather than the automatic processes in handling visuospatial representations (Andersen, Snyder,

Bradley, & Xing, 1997) and current spatial state of the body (Wolpert, Goodbody, et al., 1998).

2.1.2.1.2 Language acquisition and communication

In humans, functional brain-imaging studies have revealed that MNS activity for grasping is closely located to Broca's area that appears to be associated with speech production, suggesting thus that human language and speech evolution may be related to MNS mechanisms (Giacomo Rizzolatti & Arbib, 1998). Interestingly, a recent fMRI study focusing gestural communication between a performer and an observer revealed that the moment-to-moment MNS activity of the observer resonated with the neural activity pattern measured in the MNS of the performer (Schippers, Roebroeck, Renken, Nanetti, & Keysers, 2010). This result supports that the information such as motor planning and action intentions may flow across brains (i.e., MNSs) during social interactions.

2.1.2.1.3 Social interaction and communication

Several studies have focused on the relationship between the MNS and autism spectrum disorders (ASD), because the ASD group typically has difficulty in social interactions and communication (Dapretto et al., 2006; Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2006; Oberman et al., 2005; Williams, Whiten, Suddendorf, & Perrett, 2001). Specifically, some positive correlations have been revealed between the MNS and ASD; that is, children with ASD have reduced frontal MNS activity (Dapretto et al., 2006), and adults with ASD have thinner cortical areas in the MNS (Hadjikhani et al., 2006). These findings suggest that MNS dysfunction highly correlates with social and communication deficits associated with ASD. Interestingly, some MNS models have implicated a failure

in the development of the MNS as one cause of ASD (Hamilton, 2008; Southgate & Hamilton, 2008; Williams et al., 2001).

2.2 Neurophysiological evidence of visuospatial transformation system

A number of fMRI and TMS results has shown that the visual perspective had little effect on the activation level of the MNS during action observation, although their activations are significantly different in imitation (Aziz-Zadeh et al., 2002; Hétu, Mercier, Eugène, Michon, & Jackson, 2011). More recently, it has been found that the frontal MNS is activated for the first-person perspective action, but not for the third-person perspective action, whereas the parietal MNS is activated regardless of perspective (Oosterhof et al., 2012). In other words, this suggests that the egocentrically transformed visuomotor representations could play a critical role in the frontal MNS, whereas view-independent coding could be an essential features of the parietal MNS. Moreover, it has been found that the right SPL has an increase of activity during observation of actions from a first person perspective, whereas the left SPL has a relatively stronger activation with the third person perspective during action observation (Hesse et al., 2009). Taken together, these findings suggest that, during action observation, the visuospatial transformation would be mainly processed in different brain areas other than the MNS. Interestingly, numerous convergent results have suggested that one possible brain region involved in mental transformation of visuospatial information would be the PPC, particularly the SPL and IPS (Andersen, 1987; Buneo & Andersen, 2006; Culham & Kanwisher, 2001; Gauthier et al., 2002; Grefkes & Fink, 2005; Wolpert, Goodbody, et al., 1998). Specifically, the SPL and IPS perform complex visuospatial transformations such as rotation and scaling

(Andersen, 1987; Buneo & Andersen, 2006; Grefkes & Fink, 2005; G. Rizzolatti, Luppino, & Matelli, 1998) by processing the input signals from the visual area V5, also called middle temporal[2] (MT), which provides selective visual motion information such as direction and speed of motion (Adelson & Movshon, 1982; Tootell et al., 1995). Moreover, it has been demonstrated that the SPL and IPS are strongly associated with spatial re-orientation requiring higher cognitive processes (Corbetta et al., 1998; Thiel et al., 2004).

2.3 Internal model framework

An internal model has been traditionally defined as a postulated neural system that simulates the behavior of the sensorimotor system interacting with the external environment (Kawato, 1999). It allows the central nervous system (CNS) to determine the most appropriate motor commands necessary to achieve desired specific movements as well as to predict the consequences of those motor commands (Kawato, 1999). Typically, two types of internal models can be considered; one is the forward model and the other is the inverse model.

2.3.1 Forward model

A forward model describes the causal process that transforms the motor commands into the sensory consequences of the corresponding actions given the current state. The notion

[2] Considering its functional roles and information flow across brain regions, it seems to be reasonable that the view-dependent mirror properties of the IFG are derived from the MT, which projects to the CB (Kujala et al., 2007) as well as to the STS, IPL, and IFG (i.e., all MNS components) (Andersen, 1987; Culham & Kanwisher, 2001; Grefkes & Fink, 2005; G. Rizzolatti et al., 1998).

of forward model could be illustrated in the following example of a reaching and grasping task. The CNS incorporating a forward model of the arm receives, as three inputs, both the current position and velocity of the arm, and also the ongoing neural command to achieve the action. Based on these inputs, the forward model can predict the future sensory consequences (e.g., changes in position of the arm) even before the neural command reaches the periphery (i.e., arm muscles). Consequently, the forward model can accurately mimic the musculoskeletal system of arm, which transforms the motor commands into the position and velocity of arm through the biomechanical system of the arm.

2.3.1.1 Neural substrates for forward model

Many behavioral and neuroimaging studies have suggested that the CB (Brandauer, Timmann, Häusler, & Hermsdörfer, 2010; R. C. Miall, Reckess, & Imamizu, 2001; Nowak, Topka, Timmann, Boecker, & Hermsdörfer, 2007; Wolpert, Miall, & Kawato, 1998) and the PPC (Desmurget et al., 1999; Mulliken, Musallam, & Andersen, 2008; Sirigu et al., 1996) could be two plausible brain structures that incorporate forward model, since both are found to be involved in sensory prediction of motor control. For example, it was revealed that patients with degenerative cerebellar damage have impairments in predictive mechanisms (i.e., forward model) (Brandauer et al., 2010). Another study showed that, by investigating the PPC activity of monkeys during arm movement, the neuronal dynamics of the PPC correlate with arm movement in a joint angle (Mulliken et al., 2008). Therefore, these results indicate that the CB and the PPC could embed forward models for sensorimotor performance such as arm movements.

These findings have been applied to various models implementing action execution, in which the fronto-parietal pathway as well as the fronto-cerebellar and in turn cerebellar-parietal pathway commonly serve as the forward model (Iacoboni et al., 1999; R. Christopher Miall, 2003). Moreover, it has been suggested that the forward computation in the parietal cortex (particularly in the IPL) must be adaptive to permanently update its content for unbiased predictions (Tani, Nishimoto, & Paine, 2008; Wolpert, 1997). Specifically, the sensory prediction errors result from a mismatch between the actual (e.g., actual position of the limb) and the predicted (e.g., predicted position of the limb) sensory consequences of the movement. The forward model is then adapted for improved future movement performance by the resulting prediction errors (Mazzoni & Krakauer, 2006; Wolpert, 1997).

2.3.2 Inverse model

The causal flow of the motor system associated to the forward model can be inverted through an inverse model that provides the motor command to achieve a certain desired result such as position and velocity (Wolpert, 1997). Still considering the previous reaching and grasping example, when the individual aims to reach the object, the desired movements (e.g., position of the object or desired trajectory for the task) are transformed into the required neural commands to perform the action.

2.3.2.1 Neural substrates for inverse model

It has been suggested that CB could be a possible brain region that is responsible for inverse model. This idea has been supported by various experimental studies such as behavior tests in cerebellar patients (Maschke, Gomez, Ebner, & Konczak, 2003) and

fMRI tests (Diedrichsen, Criscimagna-Hemminger, & Shadmehr, 2007). For example, Maschke et al. (2003) have discovered that cerebellar patients fail to adapt their motor commands during arm movement that is subjected to an unknown mechanical perturbation (i.e., force field). Besides the CB, it has been also demonstrated that the premotor cortex plays an important role in the inverse dynamics (Kawato & Gomi, 1992; R. Christopher Miall, 2003; Tani et al., 2008) and inverse kinematics (Seitz et al., 1994). In particular, Seitz et al. (1994) showed in their PET study that the CB and the premotor cortex engage in learning of hand trajectories, which shows that these two structures are responsible for the inverse kinematics. Moreover, it was revealed that the CB and the premotor cortex increase in neural activity during frequent visual feedback, which indicates that these areas form a feedback network (Vaillancourt, Mayka, & Corcos, 2006). Consequently, this finding indirectly supports that these two areas are potential brain areas to incorporate inverse models, because the peripheral feedback signal, which is based on the difference between the actual and desired movements, is typically used to generate the appropriate motor command by the inverse model (R. J. Gentili et al., 2015; Gomi & Kawato, 1992; Guenther & Ghosh, 2003; Imamizu et al., 2000).

Furthermore, several modeling studies have proposed detailed neural models of the CB pathways that, after a learning period, are able to compute the inverse dynamics. This suggests that the CB can implement inverse models to control planar arm movements in horizontal and vertical workspaces (Ebadzadeh, Tondu, & Darlot, 2005; R. J. Gentili et al., 2009; Schweighofer, Arbib, & Kawato, 1998; Spoelstra, Schweighofer, & Arbib, 2000). Subsequent researches including the premotor and motor cortices have also suggested that these neural structures could implement the inverse kinematics (Bullock,

Grossberg, & Guenther, 1993; R. J. Gentili et al., 2015; R. J. Gentili, Oh, Molina, & Contreras-Vidal, 2011; Guenther & Barreca, 1997; Oh, Gentili, Reggia, & Contreras-Vidal, 2011; Oh et al., 2012; Schweighofer et al., 1998; Vilaplana & Coronado, 2006).

Generally, a crucial problem related to the inverse computation lies in the fact that there is not a unique solution, thus the same movement can be produced employing an infinite number of combinations of parameters (e.g., stiffness, angular configuration). For instance, when considering the upper limb, a set of angles determines a unique end-point position (i.e., forward computation). On the other hand, when considering a unique end-point position, the inverse computation can result in various sets of angular configurations. However, if the inverse model has been appropriately learned, then the actual and desired movements should be same (Wolpert, Miall, et al., 1998).

2.3.3 Combination of forward and inverse models

As stated above, the inverse model takes as input the desired position and provides as output the corresponding motor command. On the other hand, the forward model takes as input the motor command and predicts the future position that will be reached. Therefore, it is crucial to note that the output of the inverse model (which is the neural command) can be provided as input to the forward model allowing both of them to interact with each other. Specifically, it has been suggested that when an inverse model generates the motor command to be sent to the musculoskeletal system, a copy of this motor command, called efference copy, is also sent to a forward model to predict the corresponding sensory consequences (e.g., changes in the position of arm). Still considering the previous example, when the individual aims to reach the target object, the sequence of desired

movement (e.g., the arm trajectory to perform) will be determined and then sent to the inverse model, which will compute the required neural motor commands triggering the arm muscles. Simultaneously, an efference copy of this motor command is sent to the forward model, which predicts the future position of the arm. Subsequently, the inverse model uses these sensory predictions for online movement control guiding the arm and hand to the target object (Desmurget et al., 1999). Consequently, the combination of two adaptive neural structures can provide a high degree of adaptability and flexibility to the brain for movement control allowing thus, adaptive interactions with novel objects and environments (Kawato, 1999). Moreover, the simulation theory suggests that the forward model is at the core of motor imagery processes for covert action execution, thus the IPL allows for mental simulation of the observed action during action observation (R. Gentili, Han, Schweighofer, & Papaxanthis, 2010; R. J. Gentili et al., 2015; Jeannerod, 2001).

2.4 Review of existing computational MNS models

2.4.1 Modular selection and identification for control (MOSAIC) model

The MOSAIC model was initially introduced in a motor control framework to provide mechanisms for decentralized automatic modular selection so as to achieve the best control for the current task (Wolpert & Kawato, 1998). The basic principle of this model is to incorporate several pairs of inverse and forward models, each of which is selected appropriately for a given environment. In this model, the activity of the forward model is often considered analogous to MNS activity due to its adaptive property during imitation. However, since the MOSAIC was initially designed in the context of pure motor control

theory without considering biological relevance, its functional components are generally not directly related to any brain structure[3] (Oztop, Kawato, & Arbib, 2006; Wolpert & Kawato, 1998). Nonetheless, the MOSAIC has been successfully extended to model action recognition and imitation processes (Haruno, Wolpert, & Kawato, 2001; Wolpert, Doya, & Kawato, 2003). Although its variations can effectively model action recognition and imitation, they have not accounted for changes in the frames of reference. In other words, the MOSAIC models do not include any relevant functional structure that is responsible for the visuospatial transformation. Therefore, the model cannot process the arbitrary actions performed in any allocentric frame of reference.

2.4.2 Demiris model

The Demiris model is proposed as an imitation architecture of primate imitation mechanisms (J. Demiris & Hayes, 2002; Y. Demiris & Johnson, 2003). In particular, the model employs a dual-route process observed in the imitation mechanisms, that is, a passive and an active architectures. Specifically, the passive process is to acquire any demonstrated movement within the capabilities of the imitator, whereas the active process is to reproduce the best movement from many possible predictions; the latter is conceptually similar to the MOSAIC model. Because the Demiris model is inspired from the MOSAIC model (particularly, the distributed forward models that learn simultaneously), it has no direct biological relevance with the MNS. Moreover, since the authors focused more on the imitation system, they assumed that two different concepts, mirror activity and imitation ability, are interchangeable (Oztop et al., 2006).

[3] Wolpert & Kawato (1998) described that their MOSAIC model can be located anywhere in the brain, and the CB is a promising region for it.

Interestingly, the Demiris model includes, only at the conceptual level, the mechanism of the PPC that accounts for the transformation of the frame of reference between the demonstrator and the imitator.

2.4.3 Mental state inference (MSI) model

The MSI model is designed considering both a visual feedback control circuit, which involves the parietal and motor cortices, and a forward prediction mechanism assigned to the MNS (Oztop et al., 2005). These two systems provide an inference mechanism required for understanding other individual's intentions, thus allowing basic imitation abilities. In other words, the MSI model includes biologically plausible MNS components as well as motor control components that are employed to perform covert and overt actions. However, this model suggests only a simplified parietal model, which can extract visual features for the control of a particular action by assuming that the observed action is already egocentrically transformed. Namely, this model does not consider the problem of visuospatial transformation. Finally, another limitation is that the MSI does not explicitly incorporate the inverse computation, which is considered as an important characteristic of the MNS in terms of visual-to-motor transformation.

2.4.4 Lopes model

The Lopes model proposes a general architecture for action imitation involving the viewpoint transformation that performs a rotation to align the demonstrator's body to that of the imitator (Lopes & Santos-Victor, 2005). Although this model is based on the biological relevance of the monkey mirror property, it emphasizes only the F5 mirror neurons without considering the PF mirror neurons. In other words, Lopes & Santos-

Victor (2005) pay attention only to a counterpart of the frontal MNS but disregard an equivalent part of the parietal MNS in their model. Moreover, this model implements the visuospatial transformation without considering any biological relevance of the PPC, and furthermore, its visuospatial transformation handles not the whole body but only the arm. Indeed the viewpoint transformation is employed with transformation matrix (i.e., a pure mathematical method) in the model, resulting thus in the absence of examination of the corresponding neural processes related to the PPC. Finally, the Lopes model interestingly includes the canonical neurons found in area F5 (see section 2.1.2) to simulate its functional feature in the object manipulation during imitation.

2.4.5 MNS and MNS2 models

The MNS/MNS2 is a system level model of monkey mirror neurons to address data on mirror neurons for grasping (J. Bonaiuto & Arbib, 2010; J. B. Bonaiuto et al., 2007; Oztop & Arbib, 2002). These models include both F5 and PF mirror neurons and validates their activation at the neurophysiological level (e.g., firing). In particular, these models can simulate behavioral data for grasping with audiovisual mirror neurons as well as mirror responses to grasp a hidden object, which was formerly visible but is currently hidden in the end state. However, the MNS/MNS2 models are designed to address the development of mirror neurons as well as neural firing patterns instead of predicting a motor control role for mirror neurons (Oztop et al., 2006). Therefore, these models do not provide internal model mechanisms to simulate the basic imitation ability nor the visuospatial transformation components that would support the mirror neurons. Therefore, this model is not able to model the imitation through observational learning.

2.4.6 Summary

In conclusion, all these computational models mentioned above are based on particular mechanisms to address behavioral and neurophysiological data on mirror systems during specific actions. However, they have not examined the following three notions; first, the important functional role of inverse computation in the frontal MNS, which transforms visual representations to motor representations; second, how two modular components of the mirror system (i.e., the IFG and IPL) response within the parieto-frontal network in real-time during observational learning for imitation and reproduction to imitate; and third, the visuospatial transformation (i.e., in the SPL/IPS) and the visual motion information processing (i.e., in the MT) that respectively provide the view-independent and view-dependent representations of the observed actions to the MNS. Particularly, it must be noted that although the SPL/IPS is not a part of the MNS per se, it is assumed to subserve the MNS to enable action imitation independently of the differences in anthropometry, distance, and visual perspective between a demonstrator and an imitator (Oh et al., 2012). Therefore, it is important to develop a neurophysiologically plausible MNS model that captures fronto-parietal dynamics in conjunction with the visuospatial transformation processes to examine the functional relationships between the IFG (for inverse computations), IPL (for sensorimotor predictions), MT (for view-dependent representation), and SPL/IPS (for view-independent representation).

2.5 Backgrounds of functional neuroimaging modeling

Prior to the development of functional brain imaging techniques, the neural correlates of human cognitive, emotional, and sensorimotor functions were mainly inferred by

examining the effects of a selective brain injury or lesion, and by studying the neuronal activities from implanted electrodes in particular brain regions (Barry Horwitz, Tagamets, & McIntosh, 1999). The revolutionary changes in the understanding of the neural basis and functioning of the human brain in response to specific stimuli have been driven by the use of various noninvasive neuroimaging techniques such as TMS, computed tomography, PET, magnetic resonance imaging (MRI), EEG, or MEG, and near-infrared spectroscopy. With these techniques, functional brain activities could be recorded in awake human subjects as they perform specific cognitive, emotional, and sensorimotor tasks. Consequently, both brain functions and the interactions between brain regions associated with specific tasks have been statistically assessed. More specifically, traditional neuroimaging studies have primarily focused on localizing neural activities to determine brain functions associated with specific tasks (e.g., Cox, 1996; Friston et al., 1995; Gold et al., 1998). In addition to these studies, recent neuroimaging researches have involved functional connectivity analyses, which examine the underlying interregional neural interactions during particular tasks or in the resting brain (e.g., Bullmore & Sporns, 2009; Greicius, Krasnow, Reiss, & Menon, 2003).

Although these statistical assessments have been generally applied to functional neuroimaging data for decades, novel approaches employing computational neural modeling techniques have been recently proposed (e.g., M. A. Arbib, Bischoff, Fagg, & Grafton, 1994; Michael A. Arbib, Billard, Iacoboni, & Oztop, 2000; Michael A. Arbib, Fagg, & Grafton, 2002; Horwitz et al., 1999; Horwitz & Tagamets, 1999; Horwitz, 2004; McIntosh et al., 1994; Nunez, 1989). These innovative neural modeling methods can fall into two categories: systems-level neural modeling and large-scale neural modeling

(Barry Horwitz et al., 1999). The systems-level neural modeling aims at assessing the task-dependent quantitative strengths of interregional brain interactions using covariance structural equation modeling (McIntosh et al., 1994). On the other hand, the large-scale neural modeling is typically used to relate synthetic neural activity simulated from large-scale neural models to actual neural activity obtained from functional neuroimaging during specific tasks (M. A. Arbib et al., 1994). Therefore, it has been called 'synthetic functional brain imaging' and is employed in the current study to validate the proposed MNS model. Moreover, this is a quantitative way to illustrate the dynamics of specific neural systems during particular cognitive or sensorimotor tasks.

2.5.1 Basic principles of functional neuroimaging techniques

According to the SA4, the MNS model developed in the present work is validated by employing a synthetic functional brain imaging model based on MRI. Therefore, this section provides an introductory overview of the principles in this imaging technique for facilitating the understanding of the generation of synthetic neuroimaging signals. In particular, this section introduces PET prior to fMRI since the underlying principles of PET are similar to those of fMRI in terms of indirect measures of cerebral blood flow (CBF) changes (B. Horwitz, Friston, & Taylor, 2000). As a natural consequence, a synthetic fMRI technique is similar to corresponding a synthetic PET technique, because the former is actually an extension of the latter (Michael A. Arbib et al., 2000; Barry Horwitz & Tagamets, 1999).

2.5.1.1 PET

PET is a nuclear imaging technique that traces the gamma radioactivity when the labeled compound (or radiotracer) accumulates in specific brain regions (Ollinger & Fessler, 1997). The radiotracers are typically isotopes with short half-lives (e.g., oxygen-15 with 122.24 s and fluorine-18 with 109.77 min), which are incorporated into a biologically active molecule such as fluorodeoxyglucose by chemical synthesis in the cyclotron machine. These synthesized radiotracers are injected into the bloodstream and taken up by active neurons. As a result, PET measures regional glucose or oxygen metabolic changes according to the type of radioisotope that is used, which reflect the amount of neuronal activity associated with specific tasks (Nasrallah & Dubroff, 2013). Moreover, PET has been used to quantify regional CBF (rCBF) with oxygen-15 labeled water. The key hypothesis of this measure is that changes in rCBF are correlated with changes in regional neuronal activity in that increases in neuronal activity entail more supplies of metabolic fuels and the increased excretion of metabolic byproducts through increased rCBF (Crosson et al., 2010). However, recently fMRI has become the preferred method because PET has very low temporal and spatial resolution.

2.5.1.2 Functional MRI

MRI technology relies on the magnetic moment (i.e., a vector quantity that determines the torque) of oxygen, which is carried by the hemoglobin (Hb) molecule in red blood cells, to measure hemodynamic responses (HRs) indirectly related to neuronal activity. Briefly speaking, the MR scanner uses radiofrequency (RF) pulses to detect the MR signals, and three mutually orthogonal magnetic field gradients to localize the MR signals

in three spatial dimensions (Ashby, 2015). In addition, typically two types of RF pulse sequences adjusted by two variables of interest such as repetition time (TR) and echo time (TE) have been used to emphasize contrast between gray and white matter tissues or between brain tissue and cerebrospinal fluid (CSF). Specifically, so-called T1-weighting with short TR and short TE causes gray matter, white matter, and CSF respectively appear dark gray, light gray, and black. On the other hand, T2-weighting with long TR and long TE makes them appear light gray, dark gray, and white, respectively. Consequently, the scanner can produce detailed images of surface and deep brain structures along with the images of CBF as the brain functions (i.e., fMRI signals).

Although various techniques such as blood oxygenation level-dependent (BOLD), arterial spin labeling, and dynamic susceptibility contrast exist to measure fMRI signals, the most common method is BOLD contrast. In this work, only BOLD contrast fMRI is described, because a BOLD model is applied to generate simulated fMRI responses.

2.5.1.2.1 BOLD contrast fMRI

BOLD contrast is a measure of the ratio of oxygenated (oxyhemoglobin, HbO2) to deoxygenated hemoglobin (deoxyhemoglobin, HbR). These two hemoglobin molecules have different magnetic properties (Ogawa, Lee, R., & W., 1990). Specifically, oxyhemoglobin is diamagnetic (i.e., repelled by the applied magnetic field), whereas deoxyhemoglobin is paramagnetic (i.e., attracted by the applied magnetic field). The change from oxyhemoglobin to deoxyhemoglobin leads to change in MR signal, so it reflects an active group of neurons at a time. Specifically, the theory is that, when a short stimulus (i.e., a stick function) is provided, an active brain area consumes more oxygen as

a result of increased oxygen metabolism (i.e., cerebral metabolic rate of oxygen, CMRO2). It triggers an increase in rCBF to deliver more oxyhemoglobin. This change causes the oxyhemoglobin-to-deoxyhemoglobin ratio (i.e., BOLD contrast) to rise above baseline to peak at approximately 5 s after stimulation that elicited these responses (Crosson et al., 2010). By the way, a momentary decrease in the BOLD response known as an 'initial dip' could be observed immediately after stimulation in a high-magnetic fMRI scanner. Following this 'peak', the BOLD signal gradually decays and decreases below prestimulation baseline, and reaches an 'undershoot' (i.e., a minimum BOLD signal intensity) around at 16 s after the end of stimulation (R. Henson & Friston, 2007). Finally, it returns to baseline over a period of 30-32 s (See Figure 1).

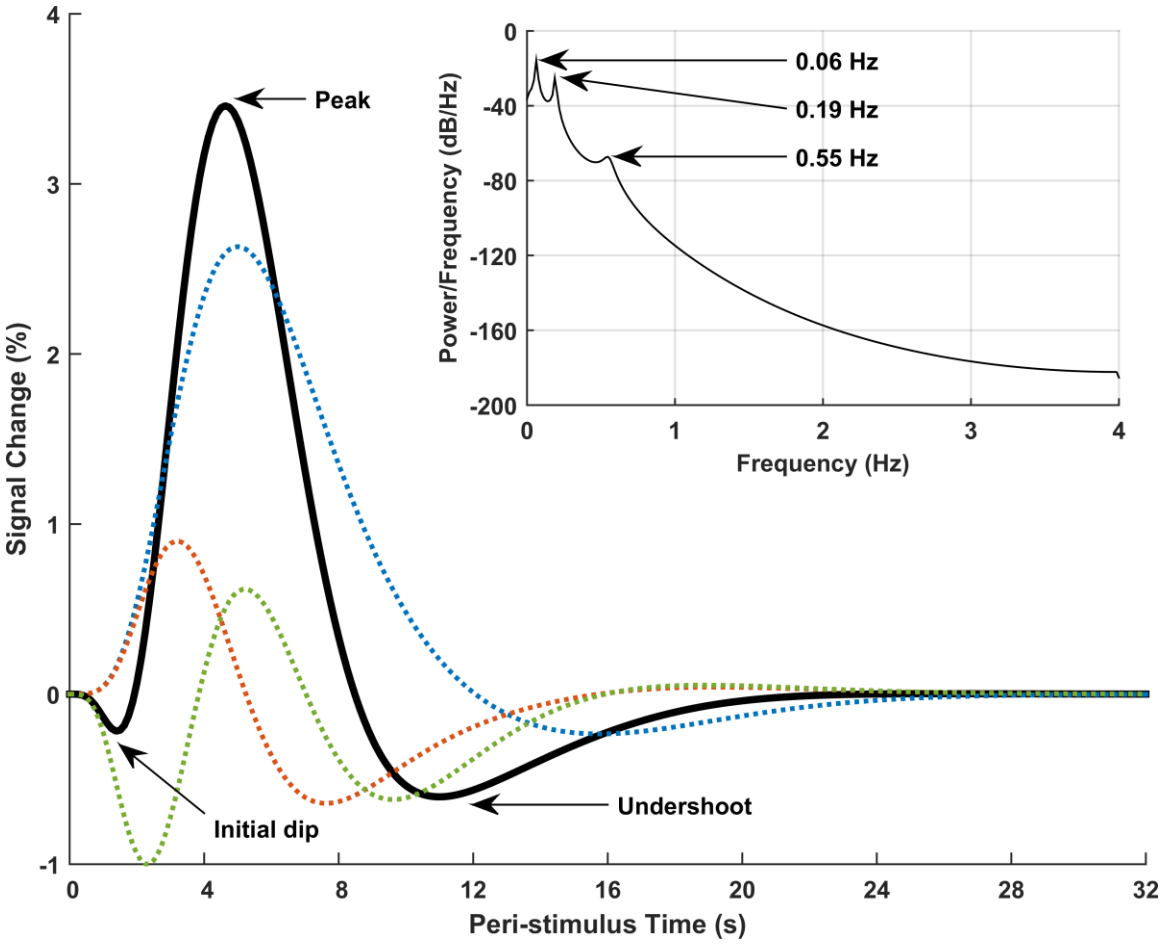


Figure 1. Typical BOLD impulse response to a stick-function shaped stimulus and its power spectrum. A typical BOLD impulse response to a stimulus presented at time 0 is expressed via a typical HRF (black solid line) generated using a linear combination of three gamma functions (i.e., the first-order Volterra kernel): the canonical HRF (blue dotted line) with its temporal (red dotted line) and dispersion (green dotted line) derivatives. Inset: Power spectral density of the typical HRF function. HRF: hemodynamic response function.

Therefore, the dynamic BOLD response depends on the changes in CBF and CMRO2 as well as venous cerebral blood volume (CBV), which are illustrated in Figure 2.

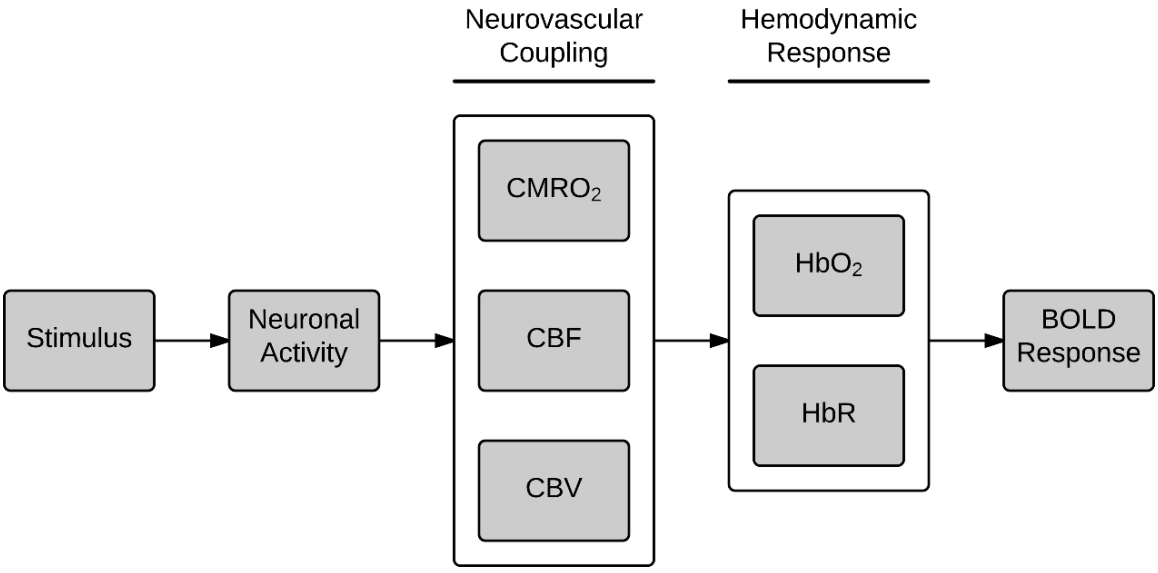


Figure 2. Schematic diagram from a stimulus to the measured BOLD response. A stimulus triggers a neuronal activity, which drives changes in neurovascular coupling (i.e., CMRO2, CBF, and CBV). These changes in neurovascular coupling lead to changes in hemodynamic response (i.e., HbO2 and HbR), resulting in the BOLD response. CMRO2: cerebral metabolic rate of oxygen; CBF: cerebral blood flow; CBV: cerebral blood volume; HbO2: oxyhemoglobin; HbR: deoxyhemoglobin.

2.5.2 Literature reviews of synthetic functional neuroimaging models

In this section, various computational models that can generate PET and fMRI data, are reviewed to reveal the relation of neural activity, CBF, and metabolism to functional neuroimaging data. The PET and fMRI based synthetic functional neuroimaging models are commonly based on complex regulatory mechanisms with numerous factors between neural activity, CBF, and metabolism. Therefore, these computational approaches allow

any neural model to effectively generate neurophysiologically plausible functional neuroimaging data under certain conditions.

2.5.2.1 PET simulation model

2.5.2.1.1 Synthetic PET model

Synthetic PET imaging was developed to use large-scale neural network models based on primate neurophysiology to predict and analyze human PET data scanned during a variety of behaviors (M. A. Arbib et al., 1994; Tagamets & Horwitz, 1998). Because both studies share the fundamental idea in terms of PET simulation, the description of the former should provide enough understanding about generation of synthetic PET. Arbib et al. (1994) used their neural network model of saccade generation (Dominey & Arbib, 1992) to simulate PET data, in which the model is composed of several brain regions as well as network connections between these structures. Particularly, in this synthetic PET study, the authors proposed the following three key hypotheses. First, each neural structure modeling the monkey brain is homologous to a region in the human brain such that their functions are same within the tasks under consideration. Second, rCBF correlates with integrated local synaptic activity in a region, and this in turn correlates with the numbers (i.e., raw PET activity) acquired in synthetic PET scans. Third, regional PET activation is computed by

$$\boldsymbol{rPET_A = \int_{t0}^{t1} \sum_B w_{B \to A}(t)\, dt} \quad (1)$$

where A is the region of interest, the sum is over all regions B that projects to A, $\boldsymbol{w_{B \to A}(t)}$ is the synaptic activity considering both firing rates of presynaptic neurons and

absolute values of both excitatory and inhibitory synapses during a time period that corresponds to the PET scan while the model performs a specific task (i.e., $\boldsymbol{firing\ rate} \times |\boldsymbol{synaptic\ strength}|$). For the comparative analysis, the subtraction paradigm is simulated, which is typically used in PET studies (B. Horwitz et al., 2000). More specifically, the change in relative synaptic activity $\boldsymbol{PET_A(1/2)}$ for region A from task 1 to task 2 is given by

$$\boldsymbol{PET_A(1/2)} = \frac{|\boldsymbol{rPET_A(1) - rPET_A(2)}|}{\boldsymbol{rPET_A(2)}} \quad (2)$$

where $\boldsymbol{rPET_A(k)}$ is obtained in Equation 1 under a task k. Arbib et al. (1994) assessed the differences in PET activity in all neural structures between two conditions using Equation 2, and showed how synthetic PET results are informative to predict human brain activity.

2.5.2.2 fMRI simulation model

As stated in section 2.5.1.2, only computational BOLD models are reviewed in this paper. The measureable BOLD fMRI signals are generated through neurovascular coupling induced by neuronal activity, and in turn through hemodynamic responses (see Figure 2). It must be noted that although the Balloon model (Appendix A) can simulate better the BOLD responses because this model emphasizes the neurovascular coupling and hemodynamic responses, it is difficult to employ this model for a general purpose large-scale neural network model such as the current MNS model in this work. In other words,

the Balloon model has potential limitations on the extensibility and flexibility of the model, therefore this approach is not considered for the current study.

2.5.2.2.1 General linear model and Convolution model

In the convolution model (R. Henson & Friston, 2007), the BOLD signal is modeled by neuronal causes that are expressed with a hemodynamic response function (HRF). The convolution model is actually an extension (i.e., with nonlinear capability) of the general linear model (GLM; Equation 3), which aims to quantify the variation of each dependent variable in terms of a linear combination of several basis functions (K. J. Friston et al., 1995). In the fMRI literatures, the dependent variable corresponds to the observed fMRI time series of each voxel, and the basis function that is also called the predictor, explanatory variable, or regressor corresponds to time course of expected BOLD response for different task conditions.

$$y(\boldsymbol{t}) = X(\boldsymbol{t})\beta + \varepsilon(\boldsymbol{t}),\ \varepsilon(\boldsymbol{t}) \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma^2}\Sigma) \quad (3)$$

where the dependent variable $y(\boldsymbol{t})$ is the observed BOLD time series, the explanatory variable $X(\boldsymbol{t})$ represents the expected BOLD time course arising from neural activity, β is time-invariant scaling parameter, $\varepsilon(\boldsymbol{t})$ is the Gaussian white noise with its standard deviation $\boldsymbol{\sigma}$ and the noise autocorrelation Σ. The neural activity is the mean synaptic activity of a group of neurons that is caused by a sequence of experimental manipulations. However, for simplicity, most fMRI studies employ a linear time-invariant (LTI) system to model the BOLD response, in which the neural activity is typically specified as just a stimulus function (R. Henson & Friston, 2007). Under a LTI

system, a set of specified explanatory variables needed to generate the BOLD time course forms a matrix known as the design matrix, in which each column corresponds to the stimulus presented. Then, the expected BOLD response $X(\boldsymbol{t})$ in Equation 3 is computed as a linear convolution of two functions, a HRF and a stimulus pattern:

$$X(\boldsymbol{t}) = (\boldsymbol{h} * u)(\boldsymbol{t}) = \int_{\boldsymbol{0}}^{\boldsymbol{T}} h(\boldsymbol{\tau})u(\boldsymbol{t} - \boldsymbol{\tau})\, \boldsymbol{d\tau} \quad (4)$$

where the impulse response $\boldsymbol{h}(\boldsymbol{t})$ is a HRF that is usually represented with the canonical HRF, the input signal $u(\boldsymbol{t})$ is the stimulus function that is usually a stick or boxcar function encoding the occurrence of an event, and $\boldsymbol{\tau}$ indexes the peri-stimulus time over which the BOLD impulse response is expressed. As an example of this process shown in Equation 4, the result of hemodynamic convolution with random events is shown in Figure 3.

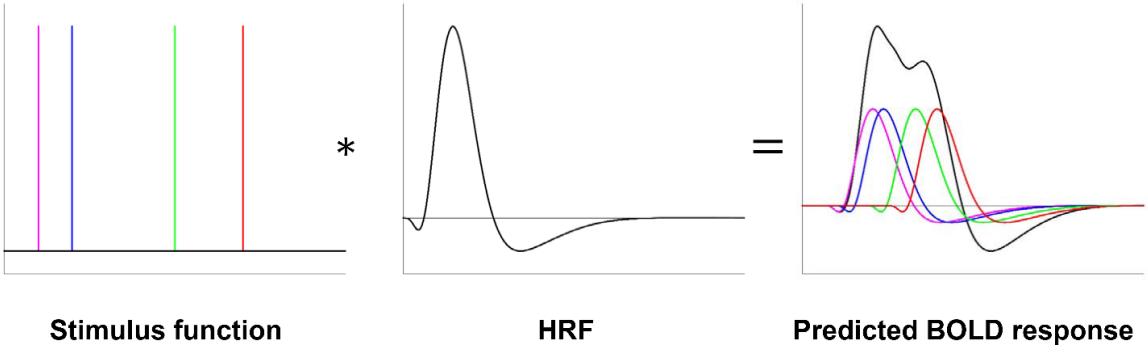


Figure 3. LTI convolution model. The BOLD signal is predicted by a linear convolution of four randomly presented events (with different colors) with a canonical HRF. Each BOLD response to

successive stimuli is linearly superposed, and the predicted BOLD response is represented with black solid line.

Recently, some studies use not only the canonical HRF, but also its temporal and dispersion derivatives for better prediction of BOLD responses. With two derivatives, the model can fit responses that are shifted in time or have extended activation durations, respectively (R. Henson & Friston, 2007). Moreover, numerous studies have assessed different types of temporal basis functions to accommodate the variability in HRF shape over brain regions as well as over individuals; for example, other popular basis functions are the finite impulse response (FIR), gamma functions, and even nonlinear convolution model using Volterra series. (K. J. Friston, Mechelli, Turner, & Price, 2000; Karl J. Friston, Josephs, Rees, & Turner, 1998; R. Henson & Friston, 2007). Specifically, two representative linear convolution models have particular properties such that FIR is a flexible basis allowing the least assumptions about the shape of HRF (Glover, 1999), whereas a set of gamma functions is the simplest basis set by sacrificing the model's degrees of freedom (R. Henson & Friston, 2007). On the other hand, the nonlinear Volterra kernels (See Appendix B) estimated from the Volterra series, which can be regarded as a Taylor series with memory capacity for dynamic systems such that the output depends on the current and past inputs, are complicated but the most powerful sets that can simulate the HR to any temporal pattern of stimuli presentations (Karl J. Friston et al., 1998). However, it was tested that the canonical HRF with its two partial derivatives, which is equivalent to the first-order Volterra kernel, are sufficient to capture

the majority of regional and experimental variabilities (Richard Henson, Rugg, & Friston, 2001).

2.5.2.2.2 Synthetic fMRI model

Two synthetic fMRI models (Michael A. Arbib et al., 2000; Barry Horwitz & Tagamets, 1999) are proposed as an extension of their synthetic PET models (M. A. Arbib et al., 1994; Tagamets & Horwitz, 1998). Particularly, Horwitz & Tagamets (1999) simulated a delayed match-to-sample task in this study with the following main hypotheses. First, BOLD signal is correlated with changes in CBF and CBV, but other minor factors are negligible. Second, BOLD signal is proportional to changes in the local field potential (LFP), which represent neural network activity at population level. Third, CBV does not alter the relation between BOLD signal and regional neural activity. Lastly, inhibitory synaptic activity, which can lead to decreased neuronal spiking and LFP, results in increased BOLD activity.

Considering the slice acquisition time for fMRI scanners, Horwitz & Tagamets (1999) reduced the integration time period represented in Equation 1 to 50 ms (when the model time step is 5 ms) instead of integrating the synaptic activity over the entire task condition. The resulting time series for each region (i.e., rCBF) is then convolved with the HRF, which is represented with a Poisson function, and noise is not added.

Specifically, the HR delay (about 5-8 s) is characterized by a Poisson function with the parameter λ in Equation 5:

$$\boldsymbol{p}(\boldsymbol{x};\boldsymbol{\lambda}) = \frac{\boldsymbol{e}^{-\boldsymbol{\lambda}}\boldsymbol{\lambda}^{\boldsymbol{x}}}{\boldsymbol{x}!}, \text{ for } \boldsymbol{x} = \mathbf{0}, \mathbf{1}, \mathbf{2}, \cdots, \tag{5}$$

where $\boldsymbol{\lambda}$ is defined as 2×T (T is the stimulus duration in seconds). The fMRI data is then obtained by sampling the hemodynamically convolved time series every TR (i.e., repetition time). This study focused on examining an event-related fMRI design, which is generally used for simple sensory or motor tasks. It must be noted that, in higher cognitive tasks, the event-related fMRI designs often cause more noisy fMRI signals, because there can be extensive task-unrelated neural activity in multiple brain regions.

Chapter 3: Methods

3.1 Mechanism of conceptual MNS model

In the present research, a new MNS model is developed by extending the conceptual MNS model that was previously designed by employing the internal model frameworks (Iacoboni et al., 1999; R. Christopher Miall, 2003). Specifically, the Miall model has been expanded by including the rostral part of the prefrontal cortex (rPFC) to trigger the intentions to imitate (Burgess, Dumontheil, & Gilbert, 2007; Dove, Pollmann, Schubert, Wiggins, & von Cramon, 2000; Meyer et al., 1997; Rogers et al., 1998) and more importantly by including the SPL and the IPS (Andersen, 1987; Buneo & Andersen, 2006; Grefkes & Fink, 2005; G. Rizzolatti et al., 1998) to implement the visuospatial transformation mechanism that provides the MNS the view-independent sensorimotor information (Figure 4). Simultaneously, the model can also process the view-dependent visual motion information through the MT (Adelson & Movshon, 1982; Tootell et al., 1995). Thus, the novelty of the proposed model is to investigate the dynamics between the frontal (i.e., adaptive inverse model) and parietal (i.e., adaptive forward scheme) MNSs as well as the view-independent and view-dependent sensorimotor processes during observational learning and imitation (Oh et al., 2011, 2012; Tani et al., 2008).

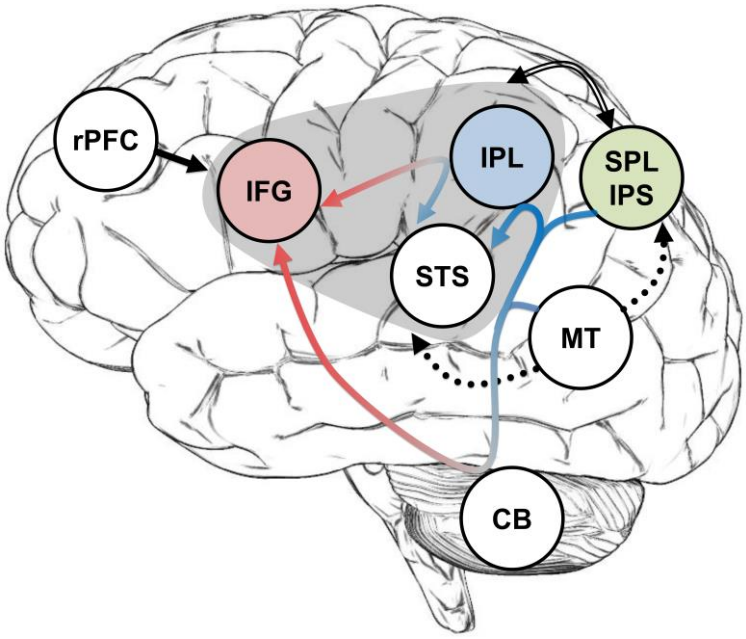


Figure 4. Conceptual MNS model overview. The model is based on the internal model framework incorporating the MT and SPL/IPS. For the sake of clarity, all the connections with the MNS (i.e., IFG, IPL, and STS) are shown with a single arrow from and to the gray shadow MNS group. For example, both the SPL/IPS and MT are actually connected with all the MNS components. The imitator either observes or executes an action based on a specific goal or intention, which is provided by the rPFC. Although it is not depicted, the MT and SPL/IPS accepts information from the visual cortex V1 to V4. The MT selectively processes the view-dependent visual motion information (dotted arrow), and provides this information to the MNS as well as the SPL/IPS. At the same time, the SPL/IPS provides the view-independent visuospatial representation (double-line arrows) to the MNS by the combination of rotation, scaling, and translation transformation. The MNS performs the inverse computation (red arrows) through the STS-IPL-IFG pathway as well as the forward computation (blue arrows) through the reverse pathway as proposed by Iacoboni et al. (1999). Moreover, another inverse and forward computations as suggested by Miall (2003) are performed through STS-IPL-CB-IFG connections with feedback signals from the MT and SPL/IPS. The SPL/IPS, IFG, and IPL are currently implemented using artificial neural networks, and the rPFC, STS, and MT is implemented in a basic level.

Imitation in this new MNS model can be accomplished through a two-phase learning process combining learning by observation and learning by execution, where the latter is voluntarily triggered by the rPFC as stated above. In the sections below these two phases are described in more details for a simple reaching and grasping task.

3.1.1 Learning by observation phase (or observational learning for imitation)

During the learning by observation phase, the imitator observes the reaching and grasping action performed by the demonstrator in the allocentric frame of reference. The visual representation encoding the observed action is sent to the MT, which is responsible for processing the selective visual motion information such as motion direction and velocity from the perceived action. When considering the vector properties of these representations, it is obvious that the visual motion information must be the view-dependent. Such a view-dependent information is relayed directly to the MNS network (i.e., IFG, IPL, and STS) and also to the SPL/IPS for further complex visuospatial processing. In particular, the SPL/IPS transforms the view-dependent visual motion information (from the visual cortex V1 to V4 as well as the MT) into the view-independent visuospatial information, which is represented in the imitator's own egocentric frame of reference.

Within the MNS network, each component employs both view-independent and view-dependent representation of the action for further processing. First, the STS examines if the observed action is already in the repertoire or unknown action, where the STS responds only to the familiar motion of specific body parts (e.g., arms or hands, but those are somewhat independent of specific states such as right arm or index finger). In

particular, the observational learning process is triggered for unseen action, otherwise the imitation process is triggered for known action, which is explained in the next section. For unseen action, the IFG is trained by mapping from the visual to motor representations (i.e., inverse computation), which is then employed to imitate the observed action. Although no movement is performed during observation, an efference copy of the motor plan is still available and is sent to the IPL that would generates the predicted sensory consequences of the corresponding action (i.e., forward computation). Meanwhile, the CB also provides the prediction error for the IFG and the IPL to adjust their internal models. At this stage, the STS inspects an exact match for the expected sensory consequences of the action and the corresponding observed action (Iacoboni, 2005). If the match fails due to a large error, the representation of the imitated action is corrected until the error decreases below a certain threshold; in other words, the observational learning is continued. Finally, the learned action can be initiated for imitation with a reasonable chance of success when the match is successful (see section 3.2.4, Figure 7-Learning by Observation panel).

3.1.2 Learning by execution phase (or imitation learning of observed action)

The observational learning described in the previous section is followed by the learning by execution, during which the rPFC triggers the regions required to imitate the action previously observed when the intention to imitate is present (Decety et al., 1997). Namely, after the learning by observation phase, the ensuing learning by execution phase is typically performed. During learning by execution, the overall processes related to the IFG, IPL, STS, CB, SPL/IPS and MT are equivalent to those previously described in the observational learning phase. The only difference is that, in parallel with the IFG to the

IPL pathway, neural drive is sent to the musculoskeletal system through the primary motor cortex to perform the actual self-action. Afterwards, the imitator egocentrically observes its own action so that the imitator can take the visual and somatosensory feedbacks (here, only the visual feedback is considered). The coincident feedback is applied by the identity transformation, and then it is employed to the MNS components to update the network components by means of the error between the sensory consequences of the self-action and the prior observed action. Consequently, the imitator can improve its action by adjusting the output with the error (see section 3.2.4, Figure 7-Learning by Execution panel).

3.2 Implementation of the computational MNS model

Based on the proposed conceptual model described above, a computational MNS model is implemented. For simplicity, the rPFC, CB, MT, and STS are implemented by simple numerical and conditional expressions and statements, thus have no adaptive capabilities. On the other hand, each of the other three components (i.e., SPL/IPS, IFG, and IPL) is implemented through the same artificial neural network architectures that learn their respective functional mappings. Specifically, the SPL/IPS learns to perform the visuospatial transformation including the rotation, scaling, and the translation. In addition, the IFG and IPL are trained to serve as the inverse and forward models, respectively. Moreover, imitation of an action through the observational learning is performed in each neural network by using a continuously repeating two-phase learning approach, that is, a repeat of learning by action observation, and in turn, learning by

action execution. This learning strategy is behaviorally more realistic in an ecologically valid context, because continuous repetitions of these two processes generally constitutes imitation.

3.2.1 Type and architecture of artificial neural network

In the current study, a radial basis function (RBF) network is used, which uses RBF for the transfer function in the hidden units (Broomhead & Lowe, 1988a, 1988b). The RBF network consists of three layers: an input layer, a hidden layer with a set of nonlinear RBF transfer functions, and an output layer that linearly summates the outputs of the hidden layer. The network measures the response of the $\boldsymbol{i}$th Gaussian RBF, $\boldsymbol{\phi_i}$, where $\boldsymbol{n}$ $(\boldsymbol{1 \leq i \leq n})$ is the number of RBFs (currently at most 50 RBFs per each output dimension), to an input v by the distance between the input and the $\boldsymbol{i}$th RBF center $\boldsymbol{c_i}$ as well as by the scaling factor of the $\boldsymbol{i}$th RBF width $\boldsymbol{\sigma_i}$, and performs a simple mapping, $\boldsymbol{f}: \mathbb{R}^{\boldsymbol{N}} \rightarrow \mathbb{R}^{\boldsymbol{2}}$, where $\boldsymbol{N}$ is the dimension of the input space, which is currently defined as 2 for the SPL/IPS and 4 for the IFG and IPL:

$$\begin{cases} \boldsymbol{f}(\mathrm{v}) & = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\omega_i \phi_i}(\|\mathrm{v} - \boldsymbol{c_i}\|) \\ \boldsymbol{\phi_i}(\boldsymbol{u}) & = \boldsymbol{e}^{(-\boldsymbol{0.5u^2}/\sigma_i^2)} \end{cases} \tag{6}$$

where v is the input vector, $\boldsymbol{f}$ is the output, $\|\cdot\|$ denotes the $\boldsymbol{L^2}$-norm, $\boldsymbol{\omega_i}$ is the $\boldsymbol{i}$th weight vector, which are fully connected as Equation 6 indicates. In particular, these 50 RBFs are equidistantly distributed to cover the whole space, which is normally scaled to be defined as $[-\boldsymbol{1}, \boldsymbol{1}] \times [-\boldsymbol{1}, \boldsymbol{1}]$, but their radii are optimally chosen by means of the following learning algorithm.

3.2.2 Learning algorithm

In a general RBF network, there are three parameters that can be optimally selected: the weights, RBF widths, and RBF centers. It is very important to optimally determine the RBFs along with their centers and widths, because too many RBFs (i.e., the large number of free parameters) can cause one of the most critical problems, overfitting, that occurs during network training. There are typically two main ways to avoid overfitting; the first is to explicitly limit the complexity of the network with a limited number of RBFs, and the second is to reduce the number of good parameter measurements by adding a regularization parameter.

In the current study, the RBF network is trained with supervised learning technique proposed by Orr (Orr, 1998). In particular, forward subset selection (Miller, 1984) using orthogonal least square (OLS) (Chen, Cowan, & Grant, 1991) is employed to determine an optimal subset of the available centers one by one. Moreover, generalized cross-validation (GCV) is used to define model selection criterion with additional parameters, which determine a moment of halting the selection process so that the criterion can estimate how the trained network can perform well on future for unknown inputs (Golub, Heath, & Wahba, 1979). Therefore, the actual learning is a repeat of forward subset selection using OLS on training set and GCV on validation set. More specifically, this learning procedure is initially performed on the coarsely spaced RBFs (i.e., a wider RBF

width range), and once it finds the best width range, it narrows down this value (i.e., fine RBF widths) to find the best RBF width.

3.2.2.1 Forward subset selection using OLS

In Equation 6, for the input vector $\mathrm{v} = [\mathrm{v}_{\mathbf{1}}, \mathrm{v}_{\mathbf{2}}, \cdots, \mathrm{v}_{\boldsymbol{m}}]$ of length $\boldsymbol{m}$, the solution of the RBF network can be represented as the following matrix form with the error signals $\boldsymbol{\varepsilon}$:

$$\boldsymbol{F} = \boldsymbol{\Phi W} + \boldsymbol{E} \tag{7}$$

Then, each term of Equation 7 is defined as following:

$$\begin{cases} \boldsymbol{F}(\mathrm{v}) & = [\boldsymbol{f_d}(\mathrm{v}_{\mathbf{1}}), \cdots, \boldsymbol{f_d}(\mathrm{v}_{\boldsymbol{m}})]^{\boldsymbol{T}} \\ \boldsymbol{\Phi}(\mathrm{v}) & = [\boldsymbol{\phi_1}(\mathrm{v}), \cdots, \boldsymbol{\phi_n}(\mathrm{v})] \\ & \boldsymbol{\phi_i}(\mathrm{v}) = \left[\boldsymbol{\phi_i^1}(\mathrm{v}_{\mathbf{1}}), \cdots, \boldsymbol{\phi_i^m}(\mathrm{v}_{\boldsymbol{m}})\right]^{\boldsymbol{T}}, \quad (\mathbf{1} \le \boldsymbol{i} \le \boldsymbol{n}) \\ & \boldsymbol{\phi_i^l}(\boldsymbol{v_l}) = \boldsymbol{\phi_i}(\|\boldsymbol{v_l} - \boldsymbol{c_i}\|), \quad (\mathbf{1} \le \boldsymbol{l} \le \boldsymbol{m}) \\ \boldsymbol{W} & = [\boldsymbol{\omega_1}, \cdots, \boldsymbol{\omega_n}]^{\boldsymbol{T}} \\ \boldsymbol{E}(\mathrm{v}) & = [\boldsymbol{\varepsilon}(\boldsymbol{v_1}), \cdots, \boldsymbol{\varepsilon}(\boldsymbol{v_m})]^{\boldsymbol{T}} \end{cases} \tag{8}$$

where $\boldsymbol{f_d}$ is the desired output. The OLS algorithm transforms $\boldsymbol{\Phi}$ into a set of orthogonal basis vectors through the QR decomposition so that it could measure the contribution of the individual RBF centers to the desired output energy from each basis vector (Chen et al., 1991):

$$\boldsymbol{\Phi} = \boldsymbol{QR} \tag{9}$$

where $\boldsymbol{Q}$ is an $\boldsymbol{m} \times \boldsymbol{m}$ orthogonal matrix, and $\boldsymbol{R}$ is an $\boldsymbol{m} \times \boldsymbol{n}$ upper triangular matrix. Assuming that $\boldsymbol{\Phi}$ has full column rank (i.e., $\boldsymbol{m} \ge \boldsymbol{n}$), the last $\boldsymbol{m} - \boldsymbol{n}$ rows of $\boldsymbol{R}$ are

entirely zero, which can be represented as the reduced QR decomposition, that is, a more compact form of Equation 9:

$$\boldsymbol{\Phi} = \boldsymbol{QR} = [\boldsymbol{Q_1} \quad \boldsymbol{Q_2}] \begin{bmatrix} \boldsymbol{R_1} \\ \boldsymbol{O} \end{bmatrix} = \boldsymbol{Q_1 R_1} \tag{10}$$

where $\boldsymbol{Q_1}$ is an $\boldsymbol{m \times n}$ matrix having orthogonal columns $\boldsymbol{q_i}$ $(\boldsymbol{1 \le i \le n})$, and $\boldsymbol{R_1}$ is an $\boldsymbol{n \times n}$ upper triangular matrix. Since the space spanned by $\boldsymbol{q_i}$ is identical to the space spanned by $\boldsymbol{\phi_i}$, Equation 7 can be rewritten in the following way:

$$\boldsymbol{F = Q_1 G + E} \tag{11}$$

where $\boldsymbol{G = R_1 W} = [\boldsymbol{g_1}, \cdots, \boldsymbol{g_n}]^{\boldsymbol{T}}$. Then, $\boldsymbol{G}$ can be solved by the OLS algorithm as

$$\boldsymbol{G} = [\boldsymbol{Q_1^T Q_1}]^{-\boldsymbol{1}} \boldsymbol{Q_1^T F} \tag{12}$$

or

$$\boldsymbol{g_i} = \boldsymbol{q_i^T f_d} / (\boldsymbol{q_i^T q_i}), \quad (\boldsymbol{1 \le i \le n}) \tag{13}$$

Such an orthogonal decomposition is iteratively processed using the classical Gram-Schmidt method, and updates $\boldsymbol{Q_1}$ and $\boldsymbol{R_1}$ as

$$\begin{cases} \boldsymbol{r_{ij}} & = \boldsymbol{q_i^T \phi_j} / (\boldsymbol{q_i^T q_i}) \\ \boldsymbol{q_j} & = \boldsymbol{\phi_j} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{j-1}} \boldsymbol{q_i r_{ij}} \end{cases}, \qquad (\boldsymbol{2 \le j \le n; 1 \le i < j}) \tag{14}$$

where $\boldsymbol{r_{ij}}$ is an $(\boldsymbol{i, j})$ element of $\boldsymbol{R_1}$, and $\boldsymbol{q_1 = \phi_1}$.

Assuming that $\boldsymbol{E}$ is uncorrelated with $\boldsymbol{\Phi}$, thus with $\boldsymbol{Q_1}$ as well, Equation 11 can be rewritten as

$$\frac{1}{m}\boldsymbol{F^T F} = \frac{1}{m}\sum_{i=1}^{n} \boldsymbol{g_i^2 q_i^T q_i} + \frac{1}{m}\boldsymbol{E^T E} \tag{15}$$

where $\frac{1}{m}\boldsymbol{g_i^2 q_i^T q_i}$ is the increment to the desired output, and an error reduction ratio can be defined as

$$\boldsymbol{\rho_i} = \boldsymbol{g_i^2 q_i^T q_i / (f_d{}^T f_d)} \tag{16}$$

Based on Equation 16, the optimal RBFs are selected by the vectors $\boldsymbol{q_i}$ with larger $\boldsymbol{\rho_i}$. By dividing Equation 15 by $\boldsymbol{F^T F}$, the OLS learning procedure is terminated at the $\boldsymbol{k}$th step when

$$\mathbf{1} - \sum_{l=1}^{k} \boldsymbol{\rho_l} < \boldsymbol{\tau} \tag{17}$$

where $\mathbf{0} < \boldsymbol{\tau} < \mathbf{1}$ is an error tolerance or the mean squared error (MSE). Then, the model finally contains $\boldsymbol{k}$ ($\leq \boldsymbol{n}$) RBFs.

3.2.2.2 GCV

In general, obtaining the minimum MSE of the training set is unlikely to make good estimates on unknown data, because some peculiarities (e.g., trends and noise) of the training set have biased the model towards the set. Therefore, in the present study, GCV is employed to estimate the potential MSE on unseen validation set and to optimize the

model by modifying the sum of squared errors (SSE) of the training set (Golub et al., 1979), which is typically defined in the following way using Equation 11:

$$\begin{aligned} \boldsymbol{SSE} &= \mathrm{E}^{\boldsymbol{T}}\mathrm{E} \\ &= (\boldsymbol{F}-\boldsymbol{Q_1 G})^{\boldsymbol{T}}(\boldsymbol{F}-\boldsymbol{Q_1 G}) \end{aligned} \tag{18}$$

In this case, the solution of the weights $\boldsymbol{G}$ is determined as Equation 12. However, in the GCV, this SSE is modified by adding a weight-decay (or ridge regression) term such that:

$$\begin{cases} \boldsymbol{SSE_{GCV}} &= \boldsymbol{SSE} + \boldsymbol{\lambda}\mathrm{G}^{\boldsymbol{T}}\mathrm{G} \\ &= (\boldsymbol{F}-\boldsymbol{Q_1 G})^{\boldsymbol{T}}(\boldsymbol{F}-\boldsymbol{Q_1 G}) + \boldsymbol{\lambda}\mathrm{G}^{\boldsymbol{T}}\mathrm{G} \\ \boldsymbol{MSE_{GCV}} &= \frac{\boldsymbol{1}}{\boldsymbol{m}}\boldsymbol{SSE_{GCV}} \end{cases} \tag{19}$$

where $\boldsymbol{\lambda}$ is the regularization parameter that controls the strength of the penalty. Through an effort to find the solution that minimizes the MSE, the regularized weight is obtained as:

$$\begin{aligned} \boldsymbol{G} &= [\boldsymbol{Q_1^T Q_1} + \boldsymbol{\lambda}\mathrm{I}_{\boldsymbol{n}}]^{-\boldsymbol{1}}\boldsymbol{Q_1^T F} \\ &= \mathrm{A}^{-\boldsymbol{1}}\boldsymbol{Q_1^T F} \end{aligned} \tag{20}$$

where $\mathrm{I}_{\boldsymbol{n}}$ is the $\boldsymbol{n}\times\boldsymbol{n}$ identity matrix, and $\mathrm{A}=\boldsymbol{Q_1^T Q_1} + \boldsymbol{\lambda}\mathrm{I}_{\boldsymbol{n}}$. Moreover, the regularization parameter $\boldsymbol{\lambda}$ and other relevant parameters are also derived from the GCV procedure as:

$$\begin{cases} \boldsymbol{\lambda} &= \frac{\boldsymbol{\eta}}{\boldsymbol{p-\gamma}}\left(\mathrm{E}^{\boldsymbol{T}}\mathrm{E}\right)/\left(\mathrm{G}^{\boldsymbol{T}}\mathrm{A}^{-\boldsymbol{1}}\mathrm{G}\right) \\ \boldsymbol{\eta} &= \boldsymbol{tr}\left(\mathrm{A}^{-\boldsymbol{1}} - \boldsymbol{\lambda}\left(\mathrm{A}^{-\boldsymbol{1}}\right)^{\boldsymbol{2}}\right) \\ \boldsymbol{\gamma} &= \boldsymbol{m} - \boldsymbol{\lambda}\cdot\boldsymbol{tr}\left(\mathrm{A}^{-\boldsymbol{1}}\right) \end{cases} \tag{21}$$

where $\boldsymbol{\eta}$ is a scaling factor, $\boldsymbol{p}$ is the number of validation set, and $\boldsymbol{\eta}$ is the effective number of free parameters (i.e., RBFs). Then, the GCV modifying the SSE (or MSE) of the training set can be simply represented as (Golub et al., 1979; Orr, 1998):

$$\boldsymbol{GCV = \frac{p}{(p-\gamma)^2} SSE_{GCV} = \frac{mp}{(p-\gamma)^2} MSE_{GCV}} \quad (22)$$

Specifically, the learning algorithm stops adding further RBFs to the network when the decreasing ratio of the GCV is less than $\boldsymbol{1.0 \times 10^{-4}}$ for at least 2 iterations. Finally, the algorithm uses backward elimination to selectively remove less significant RBFs that can be added to the network at the last two iterations when the OLS is employed, only if these last RBF bases are not the minimum values so far.

3.2.3 Computation in each component

3.2.3.1 rPFC

The rPFC is implemented as a simple conditional statement that switches between observational learning and imitation.

3.2.3.2 MT

The MT extracts the direction of the action, which is represented with two dimensional vectors. The first dimension is the distance $\boldsymbol{r}$ between the imitator's center point and the demonstrator's workspace where the demonstrator's end-effector is placed at each time $\boldsymbol{t}$ (Figure 5). Similarly, the second dimension is the relative angle $\boldsymbol{\varphi}$ from the imitator's

viewpoint to the demonstrator's workspace where the demonstrator's end-effector is placed at time $\boldsymbol{t}$ (Figure 5).

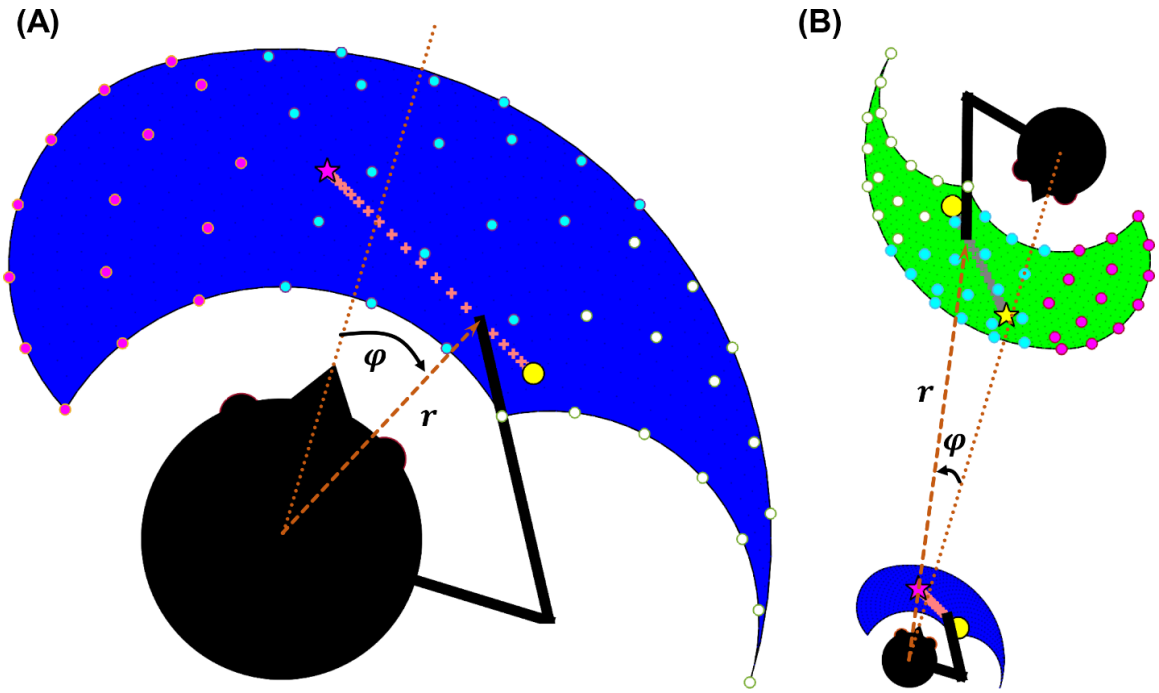


Figure 5. The relative visual motion representation of the action in the MT. (A) An imitator observes its self-action from an initial position with a yellow circle to a target magenta star. The imitator extracts the visual motion representation of the action, which is represented by a two-dimensional vector (r and φ) at each time. In particular, the vector components are roughly represented by the representative points in the left side (magenta circles), center (cyan circles), and right side (white circles) of the imitator with respect to the imitator's viewpoint. (B) At this time, the imitator (lower left side with blue workspace) observes an action (from an initial position with a yellow circle to a target with a yellow star) performed by a demonstrator (upper right side with green workspace). The visual motion representation of the action is also described by a two-dimensional vector at each time, which is roughly represented by the representative points placed relatively in the left side (white circles), center (cyan circles), and right side (magenta circles) of the demonstrator from the imitator's viewpoint.

However, this two dimensional direction of the action does not need to be precise. In practice, this vector point is approximated by the representative points relatively in the center (cyan circles), left side (white circles), and right side (magenta circles) of the imitator:

$$\boldsymbol{a}(\boldsymbol{t}; \mathbf{0} \leq \boldsymbol{t} \leq \boldsymbol{n}) = \{(\boldsymbol{r_0}, \boldsymbol{\varphi_0}), \cdots, (\boldsymbol{r_t}, \boldsymbol{\varphi_t}), \cdots (\boldsymbol{r_n}, \boldsymbol{\varphi_n})\} \quad (23)$$

where $\boldsymbol{a}$ is an observed action from time $\mathbf{0}$ to time $\boldsymbol{n}$. Therefore, the MT can provide the MNS as well as the SPL/IPS with the view-dependent directional representation of the observed action.

3.2.3.3 SPL/IPS

Three visuospatial transformation rules (i.e., rotation, translation, and scaling) are trained in the SPL/IPS using the visual motion information provided by the MT as well as visuospatial information provided by the V1 to V4. The SPL/IPS aims to help the imitator to solve and generalize the mapping $\boldsymbol{f_{VST}}: \mathbb{R}_{\boldsymbol{P}}^{\boldsymbol{2}} \rightarrow \mathbb{R}_{\boldsymbol{I}}^{\boldsymbol{2}}$, where $\mathbb{R}_{\boldsymbol{P}}^{\boldsymbol{2}}$ and $\mathbb{R}_{\boldsymbol{I}}^{\boldsymbol{2}}$ are two-dimensional workspaces respectively in the performer-centered ($\boldsymbol{P}$) and the imitator's egocentric ($\boldsymbol{I}$) frame of reference. The performer can be either a demonstrator ($\boldsymbol{D}$) or the imitator according to the condition; that is, the performer is a demonstrator during observational learning, and the performer is the imitator during imitation execution. In other words, the domain $\mathbb{R}_{\boldsymbol{P}}^{\boldsymbol{2}}$ in the mapping rule can be either $\mathbb{R}_{\boldsymbol{I}}^{\boldsymbol{2}}$ or $\mathbb{R}_{\boldsymbol{D}}^{\boldsymbol{2}}$, in which the latter is the demonstrator-centered ($\boldsymbol{D}$) frame of reference.

In general, the mapping includes various combinations of translation, rotation, scaling, and reflection where each of them can be specifically described as (Frank, 1998; Lopes & Santos-Victor, 2005):

i) *Translation*: A 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) *Rotation*: Allowing the imitator facing the same direction with the demonstrator by rotating the orientation of the axial frame of the imitator,

iii) *Scaling* and *Reflection*: So-called personalization methods for the imitator to understand the observed actions by changing the ratio and shape of the body (i.e., scaling) or the handedness (i.e., reflection).

In the current study, the reflection transformation is not considered, thus the visuospatial transformation process in the SPL/IPS is to approximate the composition of three affine transformation functions, which can be formulated in the following mathematical forms:

$$\begin{cases} \boldsymbol{f_{VST}} & = \boldsymbol{T_S}(\boldsymbol{A_I}, \boldsymbol{A_P}) \circ \boldsymbol{T_{R_z}}(\boldsymbol{\theta_I}, \boldsymbol{\theta_P}) \circ \boldsymbol{T_T}(\boldsymbol{O_I}, \boldsymbol{O_P}) \\ \boldsymbol{T_S}(\boldsymbol{u}, \boldsymbol{v}) & = \begin{pmatrix} \boldsymbol{u_x}/\boldsymbol{v_x} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{u_y}/\boldsymbol{v_y} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \\ \boldsymbol{T_{R_z}}(\boldsymbol{u}, \boldsymbol{v}) & = \begin{pmatrix} \boldsymbol{cos}(\boldsymbol{u}-\boldsymbol{v}) & -\boldsymbol{sin}(\boldsymbol{u}-\boldsymbol{v}) & \mathbf{0} \\ \boldsymbol{sin}(\boldsymbol{u}-\boldsymbol{v}) & \boldsymbol{cos}(\boldsymbol{u}-\boldsymbol{v}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \\ \boldsymbol{T_T}(\boldsymbol{u}, \boldsymbol{v}) & = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \boldsymbol{u_x}-\boldsymbol{v_x} \\ \mathbf{0} & \mathbf{1} & \boldsymbol{u_y}-\boldsymbol{v_y} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix} \end{cases} \quad (24)$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ indicate input variables, $\boldsymbol{A_I}$ and $\boldsymbol{A_P}$ are the anthropometric data, $\boldsymbol{\theta_I}$ and $\boldsymbol{\theta_P}$ are the viewpoint angles, $\boldsymbol{O_I}$ and $\boldsymbol{O_P}$ are the position vectors, $\boldsymbol{T_S}$ is the scaling matrix, $\boldsymbol{T_{R_z}}$

rotates around the $\boldsymbol{z}$ axis, $\boldsymbol{T_T}$ is the translation matrix, and finally $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the X and Y components in the Cartesian coordinates. It must be noted that the performer ($\boldsymbol{P}$) becomes either the demonstrator ($\boldsymbol{D}$) or the imitator ($\boldsymbol{I}$) according to the condition, observational learning and action execution, respectively. In particular, the rotation angle around the $\boldsymbol{z}$ axis $\boldsymbol{\theta_z}$ is the angular displacement for the mental rotation from the performer to the imitator's viewpoint (Figure 6):

$$\boldsymbol{\theta_z = \theta_I - \theta_P} \tag{25}$$

where the rotation is counterclockwise (CCW) if $\boldsymbol{0° < \theta_z \leq 180°}$, and clockwise (CW) if $\boldsymbol{-180° < \theta_z < 0°}$, respectively. Therefore, the neural network representing the rotation transformation is actually composed of two subnetworks: one for the CW network, and the other for CCW network. The consideration of these dual subnetworks is guided by neurophysiological evidence found in humans (Cohen et al., 1996).

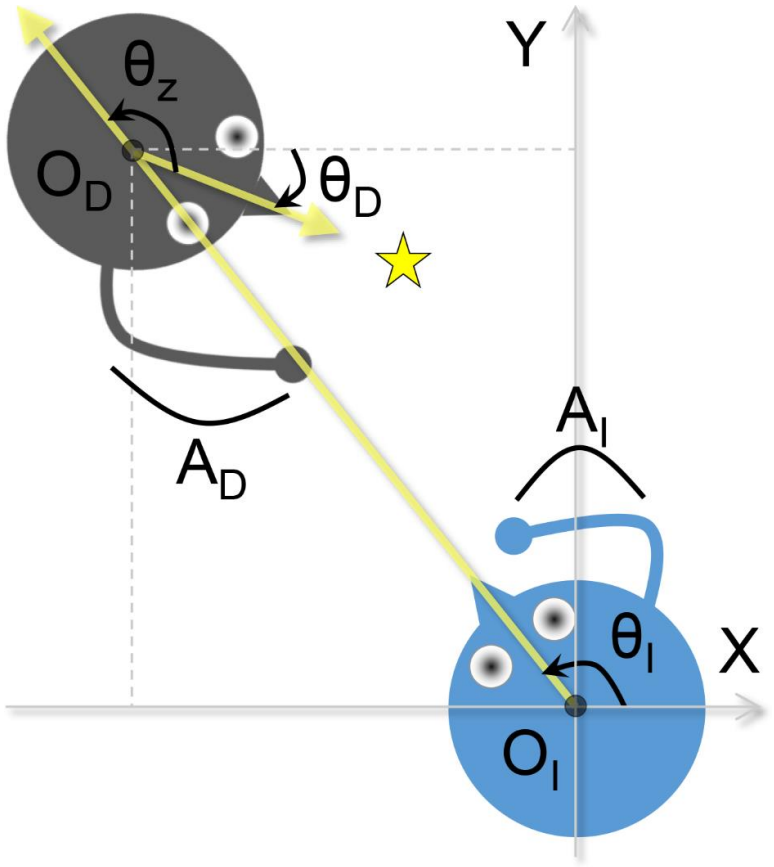


Figure 6. The visuospatial transformation in the SPL/IPS. Two agents have their own frame of reference represented by their current positions (i.e., O_I and O_D) and viewpoints (i.e., yellow lines of sight). The imitator (blue) observes reaching for a yellow star-shaped object performed by the demonstrator (gray). X and Y represent the coordinate axes of the two-dimensional global (or absolute) coordinate system. θ_I and θ_D: angles towards an imitator (I) or a demonstrator's (D) line of sight from the X-axis; θ_z: the rotation angle from the demonstrator to the imitator's viewpoint; A_I and A_D: the anthropometric data (e.g., the length of forearm).

Each of the four subnetworks (i.e., scaling, translation, CW, and CCW) is separately trained by using at most 100 RBFs (again, up to 50 RBFs per each output dimension) in the normalized workspace, $\{(\boldsymbol{x}, \boldsymbol{y})|\boldsymbol{x}, \boldsymbol{y} \in [-\mathbf{1}, +\mathbf{1}]\}$. Once these primitive transformation networks are trained, a composite network directly representing SPL/IPS response is

trained in real time for any configuration of the positions, frames of reference, and viewpoints between the demonstrator and the imitator. In general, two composite networks can be considered; one is a composite of scaling, translation, and CW networks, whereas the other is a composite of scaling, translation, and CCW networks. In practice, totally 2 (in $\boldsymbol{x}$ dimension) + 2 (in $\boldsymbol{y}$ dimension) RBFs are added for both scaling and translation networks, and 6 (in $\boldsymbol{x}$ dimension) + 14 (in $\boldsymbol{y}$ dimension) RBFs are included for both CW and CCW networks. In addition, totally 23 (in $\boldsymbol{x}$ dimension) + 24 (in $\boldsymbol{y}$ dimension) RBFs are used to represent a composite of all the subnetworks. In this case, the mean RBF width is 0.2098 and the RBF centers are uniformly distributed in each dimension when the whole workspace is normalized to a range of [-1, 1] as stated above.

3.2.3.4 STS

The STS is implemented as a simple associative memory for the template of the body parts that are associated with the observed action. For example, if an observed action (e.g., reaching) exists in the STS, the rPFC triggers the imitation execution condition to actually perform the corresponding observed action (e.g., reaching). Otherwise, the rPFC triggers the observational learning condition, in which, for the observed reaching task, a set of template information such as the end-effector part (i.e., arm), current action type (i.e., reaching), and the associated action (i.e., a directional vector from an initial point to a target point) are stored in the STS. Thus, this indicates that the observed action is now in the imitator's repertoire.

3.2.3.5 CB

The CB is implemented as a simple expression calculating two errors; one is between the observed actions and desired actions to update the frontal MNS (i.e., inverse model) during observational learning, and the other is between the actual actions perceived through sensory feedbacks and the predicted actions to update the parietal MNS (i.e., forward model) during action execution.

3.2.3.6 IFG

The IFG represents an adaptive inverse model that can be described by the mapping $\boldsymbol{f_{INV}}: \mathbb{R}_{\boldsymbol{V}}^{\boldsymbol{4}} \rightarrow \mathbb{R}_{\boldsymbol{M}}^{\boldsymbol{2}}$, where $\mathbb{R}_{\boldsymbol{V}}^{\boldsymbol{4}}$ specifies the observed action in the visual ($\boldsymbol{V}$) domain with two-dimensional view-independent visuospatial representation from the SPL/IPS and another two-dimensional view-dependent visual motor representation from the MT, and $\mathbb{R}_{\boldsymbol{M}}^{\boldsymbol{2}}$ is the two-dimensional motor plan in the motor ($\boldsymbol{M}$) domain.

Assuming the analytical forward model is formulated as the following (see Table 1 in section 3.4):

$$\begin{cases} \boldsymbol{x} & = \boldsymbol{l_1}\, \boldsymbol{cos}\, \boldsymbol{\theta_1} + \boldsymbol{l_2}\, \boldsymbol{cos}(\boldsymbol{\theta_1} + \boldsymbol{\theta_2}) \\ \boldsymbol{y} & = \boldsymbol{l_1}\, \boldsymbol{sin}\, \boldsymbol{\theta_1} + \boldsymbol{l_2}\, \boldsymbol{sin}(\boldsymbol{\theta_1} + \boldsymbol{\theta_2}) \end{cases} \tag{26}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the predicted end-effector position of the performer (i.e., either the demonstrator or the imitator) relative to the imitator that determines the physically reachable horizontal planar workspace, $\boldsymbol{l_1}$ and $\boldsymbol{l_2}$ are the length of each link (i.e., upper arm, and forearm) composing the right upper limb, and finally $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ are the joint

angle of these two links, respectively. In that case, the analytical solution of the inverse model for this forward model can be represented as following:

$$\begin{cases} \boldsymbol{\theta_1} & = \boldsymbol{tan^{-1}}(\frac{y}{x}) - \boldsymbol{tan^{-1}}\left(\frac{\sqrt{4l_1^2l_2^2-(x^2+y^2-l_1^2-l_2^2)^2}}{x^2+y^2+l_1^2-l_2^2}\right) \\ \boldsymbol{\theta_2} & = \boldsymbol{tan^{-1}}(\sqrt{(\frac{2l_1l_2}{x^2+y^2-l_1^2-l_2^2})^2-\mathbf{1}}) \\ \dot{\boldsymbol{\varphi}}_{\boldsymbol{t}} & = \frac{\mathbf{1}}{r_t}\|\mathrm{v}_{\perp}(\dot{\boldsymbol{x}},\dot{\boldsymbol{y}})\|\,\boldsymbol{sin}\,\boldsymbol{\varphi_t} \\ \mathrm{v}_{\perp}(\dot{\boldsymbol{x}},\dot{\boldsymbol{y}}) & = \begin{pmatrix} \boldsymbol{cos}\,\boldsymbol{\theta_I} & \boldsymbol{sin}\,\boldsymbol{\theta_I} \\ -\boldsymbol{sin}\,\boldsymbol{\theta_I} & \boldsymbol{cos}\,\boldsymbol{\theta_I} \end{pmatrix}\begin{pmatrix} \boldsymbol{\partial x/\partial\theta_1} & \boldsymbol{\partial x/\partial\theta_2} \\ \boldsymbol{\partial y/\partial\theta_1} & \boldsymbol{\partial y/\partial\theta_2} \end{pmatrix}\begin{pmatrix} \dot{\boldsymbol{x}} \\ \dot{\boldsymbol{y}} \end{pmatrix} \end{cases} \quad (27)$$

where the first two equations are for inverse kinematics, and the other two are for angular velocity of the end-effector with respect to the imitator's viewpoint $\boldsymbol{\theta_I}$, which is internally used in the IFG to represent the view-dependent visual motion information. Moreover, $\boldsymbol{\varphi_t}$ is the relative angle from the imitator's viewpoint to the demonstrator's workspace where the demonstrator's end-effector is placed in the at time $\boldsymbol{t}$, $\boldsymbol{r_t}$ is the distance to this point from the imitator's center point at this moment (see Equation 23), $\mathrm{v}_{\perp}$ is the perpendicular component of the motion (or angular velocity) with respect to $\boldsymbol{\varphi_t}$, and $\|\cdot\|$ denotes the $\boldsymbol{L^2}$-norm. Therefore, the adaptive inverse model is to approximate this Equation 27; however, detailed operations are slightly different in the observation and the execution phases. For example, the predicted sensory consequences of the corresponding desired action have more important role in the learning of the adaptive inverse model during observational learning because the imitator mentally simulates the observed action during this phase. On the other hand, the observed self-performed action

exerts more significant influence on training of the adaptive inverse model during actual action execution. These operations can be represented with the following expressions:

$$\begin{aligned}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) &= \boldsymbol{f_{INV}}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \boldsymbol{r_t}, \boldsymbol{\varphi_t}) \\ (\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}) &= \begin{cases} (\mathbf{1}-\boldsymbol{\delta})(\boldsymbol{x}, \boldsymbol{y}) + \boldsymbol{\delta}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}), & \text{in observational learning} \\ \boldsymbol{\delta}(\boldsymbol{x}, \boldsymbol{y}) + (\mathbf{1}-\boldsymbol{\delta})(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}), & \text{in action execution} \end{cases} \end{aligned} \quad (28)$$

where the input vector $(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}})$ is the weighted end-effector position, another input vector $(\boldsymbol{r_t}, \boldsymbol{\varphi_t})$ is the visual motion representation of the action provided by the MT. Particularly, the weighted end-effector position is calculated from the view-independent performer's end-effector position $(\boldsymbol{x}, \boldsymbol{y})$ through the SPL/IPS and the predicted end-effector position $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ through the IPL by a weighting factor $\boldsymbol{\delta}$ $(\mathbf{0.5} \leq \boldsymbol{\delta} \leq \mathbf{1})$, respectively. Generally, this weighting factor is initially set as a higher value (e.g., 0.9), but gradually decreases during learning, thus it can be simulated using a prediction error scaled between 0.5 and 1.

After training, a total of 46 (in $\boldsymbol{\theta_1}$ dimension) and 48 (in $\boldsymbol{\theta_2}$ dimension) RBFs are able to implement the inverse model, and the mean RBF width is 0.3420 and the RBF centers are uniformly distributed in the normalized space of [-1, 1].

3.2.3.7 IPL

The IPL represents an adaptive forward model that can be described by the mapping $\boldsymbol{f_{FWD}}: \mathbb{R}_{\boldsymbol{M}}^{\mathbf{4}} \rightarrow \mathbb{R}_{\boldsymbol{V}}^{\mathbf{2}}$, where $\mathbb{R}_{\boldsymbol{M}}^{\mathbf{4}}$ is the efference copy of the motor plan in the motor $(\boldsymbol{M})$ domain including two-dimensional joint angles for each link composing the right upper limb (i.e., upper arm and forearm) provided by the IFG and additional two-dimensional view-dependent visual motor representation provided by the MT, and $\mathbb{R}_{\boldsymbol{V}}^{\mathbf{2}}$ is the two-

dimensional predicted end-effector position in the visual ($\boldsymbol{V}$) domain. In practice, the adaptive forward model is to estimate the Equation 26 and the relative angular velocity formulated in the last two equations of Equation 27, which is also internally used in the IPL to represent the view-dependent visual motor representation. The IPL is trained during the actual reproduction of the action for imitation. After learning, the mean RBF width becomes 0.5590, and the RBF centers (totally 27 in $\boldsymbol{x}$ dimension plus 26 in $\boldsymbol{y}$ dimension) are uniformly distributed in the whole workspace scaled between -1 and 1.

3.2.4 Online and batch learning procedure

For network training, the imitator either observes the demonstrator's action to learn it or executes the observed action. Particularly, two different network update methods are employed for the three neural networks (i.e., SPL/IPS, IFG, and IPL): batch and online update methods. In the batch method, each network is trained separately by observing the demonstrator's action. On the other hand, in the online method, all the neural weights are updated sequentially (i.e., the SPL/IPS followed by the IFG, and again followed by the IPL) by the corresponding errors computed for each step, and this is repeated until all three networks are completely trained based on the learning rules (see 3.2.2; Figure 7). Although the batch update method provides accurate and robust results and is easier to implement when multiple neural networks are involved, it is less feasible in simulating the interaction between the co-working neural networks. Therefore, in the current study, the online update method is emphasized to train the IFG and IPL as well as the SPL/IPS as stated in section 3.2.4.

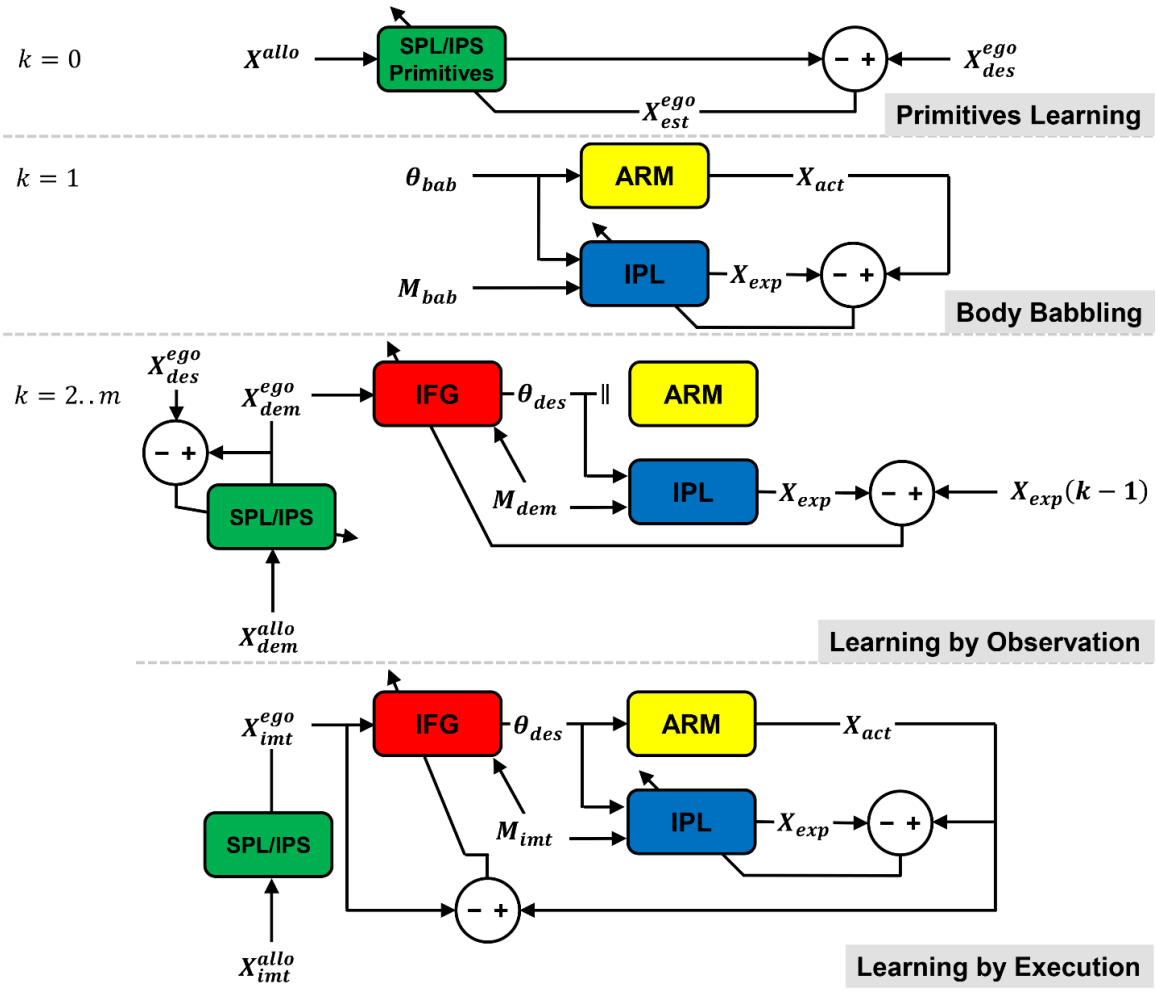


Figure 7. Overall learning procedure employing online update method. At first, four visuospatial transformation primitives of the SPL/IPS are trained in advance (k = 0). For a body babbling, the forward model of the IPL is trained for one step at k = 1. Next, the two-phase learning strategy is implemented by continuous repetitions of learning by observation and learning by execution until all three neural networks are fully trained (k = 2 to m). Particularly, in learning by observation phase, the SPL/IPS and the IFG are trained using the first equation in Equation 24 and the corresponding rule in Equation 28, respectively. During learning by execution, the IFG are updated with the corresponding rule in Equation 28 and the IPL is trained using Equation 26 and the last two equations (i.e., angular velocity of the end-effector) in Equation 27.

It must be noted that, as described in 3.2.3.3 the four primitives of the SPL/IPS (i.e., scaling, translation, CW, and CCW) are already trained (Figure 7; $\boldsymbol{k = 0}$). However, for the SPL/IPS, a composite function of these primitives is trained in the online training method (Figure 7; $\boldsymbol{k = 2..m}$). Moreover, for the online training method. a so-called 'body babbling' is implemented, in which infants practice their movements through self-generated activity (Meltzoff & Moore, 1997). In other words, there is a pre-training session of the IPL only for a step corresponding to the body babbling, which results in gaining one RBF in the IPL that produces inaccurate self-generated movements at this moment (Figure 7; $\boldsymbol{k = 1}$).

The subsequent learning is performed by using a continuous cycle of a two-phase learning process, that is, a repeat of learning through action observation followed by learning through action execution (Figure 7; $\boldsymbol{k = 2..m}$), where $\boldsymbol{m}$ is the number of iterations or added RBFs. This learning strategy is behaviorally more realistic in an ecologically valid condition, because the action must be observed (i.e., learning by observation) prior to it being actually performed (i.e., learning by action execution). Generally, the same process needs to be repeated several times for complete acquisition of the skilled action. Specifically, the SPL/IPS and IFG are trained during learning by observation phase, where the IFG and IPL are trained during learning by action execution; here, the action is already in the imitator's own frame of reference, so the SPL/IPS works as an identity transformation.

3.3 Implementation of the synthetic neuroimaging model

3.3.1 BOLD fMRI simulation

As stated in section 2.5.2.2.1, the relationship between the stimulus and task-evoked BOLD response is generally modeled using a LTI system, where the BOLD response depends on both the input signal and the corresponding HRF (see Equation 4). In this section, detailed modeling methods are described for both HRF and BOLD response, which are used to generate the BOLD fMRI signals for the proposed MNS model (see Figure 8).

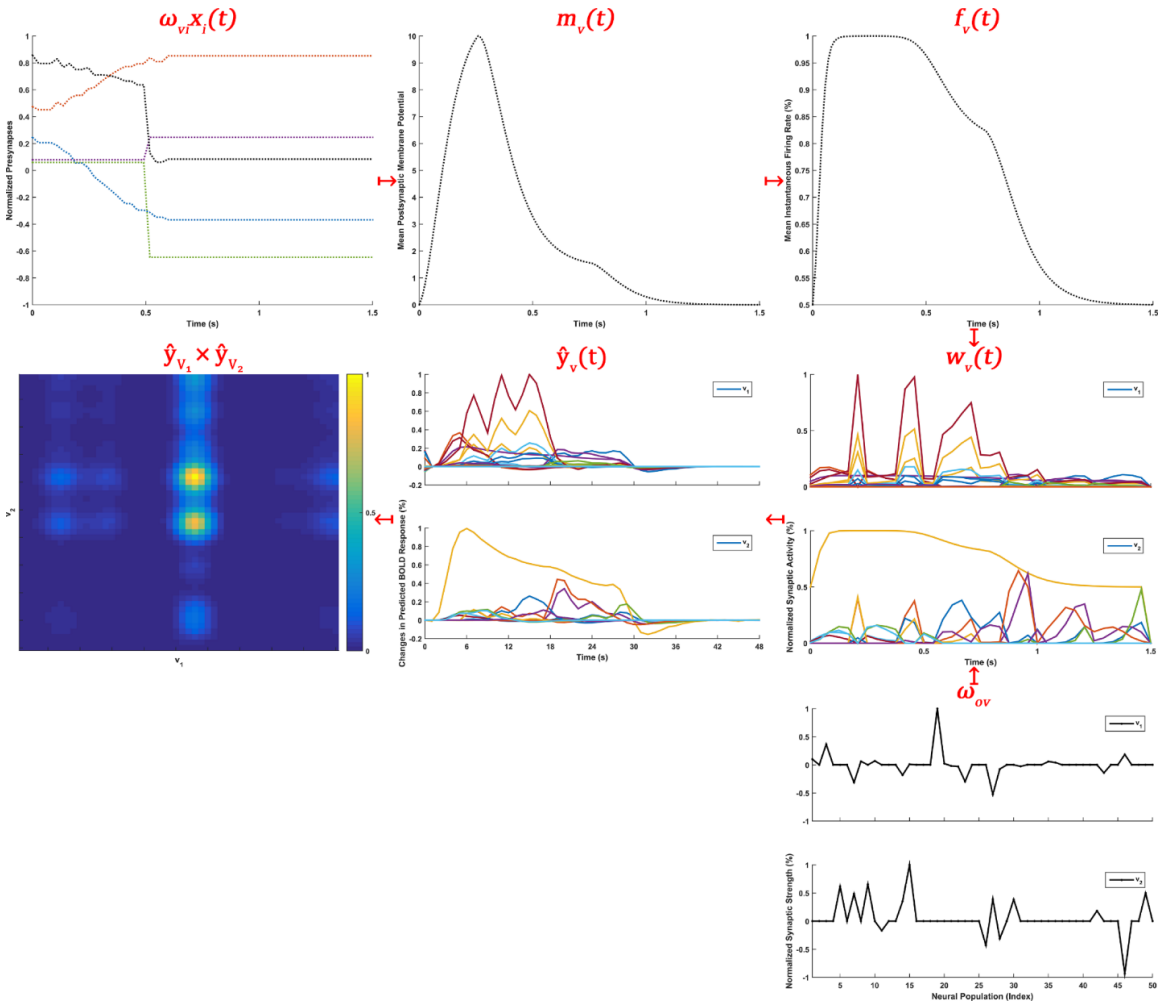


Figure 8. The synthetic BOLD fMRI generation process. Top-left panel is the membrane potential during a reach to grasp action for 1.5 s, where the four lines indicate the activity of four input vector in the neural network. Top-center panel is the average postsynaptic membrane potential derived from Equation 35. Top-right panel is the mean instantaneous population firing rate (i.e., population coding) by Equation 34. Bottom-right panel is the synaptic strength to the group of output neural population obtained through the neural network. Middle-right panel is the mean synaptic activity of a subpopulation neuron in Equation 34. Middle-center panel is the predicted BOLD response of all voxels in Equation 33. Middle-left panel is the two-dimensional representation of the FWHM filtered BOLD responses. FWHM: full width at half maximum.

3.3.2 HRF modeling

A single gamma function $\boldsymbol{g}$ has been proved to provide a good approximation to the BOLD impulse response (Boynton, Engel, Glover, & Heeger, 1996):

$$\boldsymbol{g(t) = \frac{1}{\tau(n-1)!}(t/\tau)^{n-1}e^{-(t/\tau)}} \tag{29}$$

where $\boldsymbol{t}$ is time, $\boldsymbol{\tau}$ is the time scaling, and $\boldsymbol{n}$ is a phase delay. However, due to its lack of details (e.g., undershoot and dispersion) observed in the BOLD impulse response, a set of gamma functions has been recently used (Figure 1). This can be mathematically represented by a Taylor expansion of the real BOLD impulse response $\boldsymbol{r}$ by a convolution of an expected HRF $\boldsymbol{h}$ and the Dirac delta function $\boldsymbol{\delta}$:

$$\boldsymbol{r(t) = \gamma \cdot (h * \delta)(t)} \tag{30}$$

where $\boldsymbol{\gamma}$ is the strength scaling. In practice, $\boldsymbol{h}$ can be approximated to $\boldsymbol{\hat{h}}$ with a gamma function $\boldsymbol{g}$ that is shifted by a small amount $\boldsymbol{\tau}$ in time. Therefore, the Equation 30 can be approximated as the estimated BOLD impulse response $\hat{\boldsymbol{r}}$:

$$\begin{cases} \hat{\boldsymbol{r}}(\boldsymbol{t}) & = \boldsymbol{\gamma \cdot \left(\hat{h} * \delta\right)(t) + \varepsilon(t)} \\ & = \boldsymbol{\gamma \cdot \hat{h}(t) + \varepsilon(t)} \\ & = \boldsymbol{\gamma \cdot g(t+\tau) + \varepsilon(t)} \\ & = \boldsymbol{\gamma \cdot \left(g(t) + \tau g'(t) + \frac{1}{2!}\tau^2 g''(t) + R_2\right) + \varepsilon(t)} \\ \boldsymbol{\varepsilon(t)} & \boldsymbol{\sim N\left(0, \sigma_g^2\right)} \\ \boldsymbol{R_2} & \boldsymbol{= \frac{1}{3!}\tau^3 g^{(3)}(t^*)} \end{cases} \tag{31}$$

where $\boldsymbol{\varepsilon}$ is the approximation error in normal distribution with the standard deviation $\boldsymbol{\sigma_g}$, the three functions, $\boldsymbol{g}$, $\boldsymbol{g'}$, and $\boldsymbol{g''}$, are respectively the zeroth to second derivatives of the

gamma function, and $\boldsymbol{R_2}$ is the Lagrange remainder for some $\boldsymbol{t^*} \in [\boldsymbol{t}, \boldsymbol{t+\tau}]$. Therefore, $\widehat{\boldsymbol{h}}$ is obtained as:

$$\begin{aligned} \widehat{\boldsymbol{h}}(\boldsymbol{t}) &= \boldsymbol{R_2} + \boldsymbol{g}(\boldsymbol{t}) + \boldsymbol{\tau}\boldsymbol{g}'(\boldsymbol{t}) + \frac{\mathbf{1}}{\mathbf{2!}}\boldsymbol{\tau}^{\mathbf{2}}\boldsymbol{g}''(\boldsymbol{t}) \\ &= {}^{\mathbf{0}}\boldsymbol{\beta} \cdot \mathbf{1}(\boldsymbol{t}) + {}^{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{1}} \cdot \boldsymbol{g}(\boldsymbol{t}) + {}^{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{2}} \cdot \boldsymbol{g}'(\boldsymbol{t}) + {}^{\mathbf{1}}\boldsymbol{\beta}_{\mathbf{3}} \cdot \boldsymbol{g}''(\boldsymbol{t}) \end{aligned} \tag{32}$$

where ${}^{\mathbf{0}}\boldsymbol{\beta}$ and ${}^{\mathbf{1}}\boldsymbol{\beta}_{\boldsymbol{k}}$ are the coefficients in a zeroth- and first-order Volterra kernel (see Appendix B). In this study, the Statistical Parametric Mapping (SPM) software (SPM12, Wellcome Trust Centre for Neuroimaging, University College London, London, UK) is used to estimate Equation 31 and 32 with $\boldsymbol{\tau} = \mathbf{1}$, $\boldsymbol{n} = \mathbf{6}$, and $\boldsymbol{t} \in [\mathbf{0}, \mathbf{32}]$ for $\boldsymbol{g}(\boldsymbol{t})$.

3.3.3 BOLD response modeling

The BOLD response is typically modeled using a convolution model; in particular, a GLM with a LTI system. However, the LTI system oversimplified the process by replacing the mean synaptic activity with the brief stimulation function (Equation 4). As a result, the output of this system becomes also simpler (i.e., the BOLD impulse response) than its original form, the BOLD response. In the present study, for more accurate BOLD response simulation, the predicted BOLD response is calculated from a convolution of the mean synaptic activity of an ensemble of neurons with the expected HRF. For this computation, the mean-field approximation is assumed in such a way that, in the artificial neural networks, each hidden unit reflects a subpopulation of neurons rather than a single neuron in each brain region so that the network weight from the unit correspond to the

regional mean neural population activity (i.e., LFP). This model can be expressed as similar as Equation 31:

$$\widehat{\boldsymbol{y}}_{\boldsymbol{v}}(\boldsymbol{t}) = \boldsymbol{\gamma} \cdot \left(\widehat{\boldsymbol{h}}_{\boldsymbol{v}} * \boldsymbol{w}_{\boldsymbol{v}}\right)(\boldsymbol{t}) + \boldsymbol{\varepsilon}_{\boldsymbol{v}}(\boldsymbol{t}) \tag{33}$$

where $\boldsymbol{v}$ is a subpopulation of neurons (or a voxel), $\widehat{\boldsymbol{y}}_{\boldsymbol{v}}$ is the predicted BOLD response of $\boldsymbol{v}$, $\boldsymbol{w}_{\boldsymbol{v}}$ is the mean synaptic activity of $\boldsymbol{v}$, $\widehat{\boldsymbol{h}}_{\boldsymbol{v}}$ is the estimated HRF of $\boldsymbol{v}$, and $\boldsymbol{\varepsilon}_{\boldsymbol{v}}$ is the Gaussian white noise of $\boldsymbol{v}$ (Figure 8-Middle center panel). In this study, $\boldsymbol{\gamma}$ is 1 and $\boldsymbol{\varepsilon}_{\boldsymbol{v}}$ is randomly generated for the signal-to-noise ratio (SNR) to be 65, which is the average temporal SNR measurement with 3×3×3 mm3 resolution at 3 tesla (T) (Triantafyllou et al., 2005). This allows that each hidden unit can be mapped onto each voxel of 55 mm3 (i.e., ~ 3.8×3.8×3.8 mm3) that is composed of about 5.5 million neurons (Logothetis, 2008), and that the RBF radius can represent the voxel size. Moreover, the total number of hidden units over all dimensions is bound to a specific value (e.g., 100), which is designed to reflect the finding that the maximum number of voxels in any cortical region is about 100 in terms of classification accuracy (D. D. Cox & Savoy, 2003).

The mean synaptic activity of a subpopulation neuron $\boldsymbol{w_v}$ in Equation 33 is obtained through the digitization (Figure 8-Middle right panel):

$$\begin{cases} \boldsymbol{w_v(t)} & = \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} \boldsymbol{f_v(t)} \cdot |\boldsymbol{\omega_{ov}}| \boldsymbol{dt} \qquad , \boldsymbol{t_{k-1} \le t < t_k} \\ \boldsymbol{f_v(t)} & = \boldsymbol{s}\big(\boldsymbol{m_v(t)}\big) \\ \boldsymbol{s(t)} & = (\mathbf{1} + \boldsymbol{e^{-t}})^{\mathbf{-1}} \\ \boldsymbol{m_v(t)} & = \sum_{i=1}^{n} \boldsymbol{\omega_{vi} x_i(t)} \sum_{j=1}^{\infty} \boldsymbol{\alpha}\big(\boldsymbol{t - \tau_i^j}\big) \\ \boldsymbol{\alpha(t)} & = \boldsymbol{A(t/\tau) e^{-(t/\tau)}} \end{cases} \tag{34}$$

where $\boldsymbol{N}$ is the number of samples, each of which is digitized in the time interval from $\boldsymbol{t_{k-1}}$ to $\boldsymbol{t_k}$ (in particular, $\boldsymbol{t_0}$ is 0 and $\boldsymbol{t_N}$ corresponds to the duration of the scan), $\boldsymbol{f_v}$ is the mean firing rate of $\boldsymbol{v}$ (Figure 8-Top right panel), and $\boldsymbol{\omega_{ov}}$ is the synaptic strength from $\boldsymbol{v}$ to the group of output neurons $\boldsymbol{o}$ (Figure 8-Bottom right panel). More importantly, it is assumed that $\boldsymbol{f_v}$ is computed not by the emission of single spikes but by the average rate of action potentials fired by the neighboring population as the Jansen-Rit neural mass model (Jansen & Rit, 1995). In other words, $\boldsymbol{f_v}$ actually reflects the mean instantaneous population firing rate (i.e., population coding) that is approximated by a sigmoid transfer function $\boldsymbol{s}$ of the mean membrane potential of $\boldsymbol{v}$ (i.e., $\boldsymbol{m_v}$). The average postsynaptic membrane potential $\boldsymbol{m_v}$ is calculated from all $\boldsymbol{n}$ incoming synapses from the input $\boldsymbol{i}$ to $\boldsymbol{v}$ (i.e., presynapses $\boldsymbol{\omega_{vi}}$), the incoming stimuli $\boldsymbol{x_i}$ of $\boldsymbol{i}$, and a synaptic transient function called the alpha function $\boldsymbol{\alpha}$ (Figure 8-Top left panel), where $\boldsymbol{\tau_i^j}$ represents the time of the $\boldsymbol{j}$th instantaneous fire (David & Friston, 2003; Jansen & Rit, 1995). The alpha function has two constants representing the maximum amplitude of the postsynaptic membrane

potential $\boldsymbol{A}$ and a time constant $\boldsymbol{\tau}$. In practice, $\boldsymbol{m_v}$ (Figure 8-Top center panel) can be simply computed by a convolution of $\boldsymbol{\omega_{vi}}$ and $\boldsymbol{\alpha}$ (David & Friston, 2003):

$$\begin{aligned} \boldsymbol{m_v(t)} &= \sum_{i=1}^{n} (\boldsymbol{\omega_{vi} x_i} * \boldsymbol{\alpha})(\boldsymbol{t}) \\ &= (\sum_{i=1}^{n} \boldsymbol{\omega_{vi} x_i} * \boldsymbol{\alpha})(\boldsymbol{t}) \end{aligned} \tag{35}$$

Specifically, the integration time interval (i.e., $\boldsymbol{t_k - t_{k-1}}$ in Equation 34) is set to 62.5 ms, which corresponds to the acquisition time of 32 slices when TR/TE is 2000/30 ms in the typical BOLD fMRI protocol at 3T; for example, the number of slices and the voxel size can be 33 and 3.3×3.3×3.3 mm3 when a 2-s TR, 30-ms TE, 78° flip angle, and 211-mm field of view with 64×64 matrix size in a gradient echo based echo planar imaging sequence on a 3T system such as MAGNETOM® Trio, A Tim System (Siemens Medical Solutions USA, Inc., Malvern, PA, USA) (Monaco et al., 2011). In addition, $\boldsymbol{n}$, $\boldsymbol{A}$, and $\boldsymbol{\tau}$ are taken to be 4, 1 mV and 10 ms for the sake of simplicity of the current computational model (Equation 34 and 35). The incoming stimuli $\boldsymbol{x_i}$ correspond to various internal representations of the reaching and grasping action (duration: 1.5 s) for each brain region (e.g., spatial kinematic trajectories in SPL/IPS and IFG, angular trajectories in IPL, etc.).

It must be noted that the SPM approximates $\widehat{\boldsymbol{h}}_{\boldsymbol{v}}$ at a higher temporal resolution of 62.5 ms (i.e., 16 Hz), which is $\Delta \boldsymbol{t} = \boldsymbol{TR/N_t} = \boldsymbol{2/32}$ with the number of time-bins per scan $\boldsymbol{N_t}$, to capture more information in the predicted BOLD signals (i.e., $\widehat{\boldsymbol{y}}_{\boldsymbol{v}}$ in Equation 33). Therefore, $\widehat{\boldsymbol{y}}_{\boldsymbol{v}}$ is decimated (i.e., downsampling) to produce the BOLD fMRI images once every 2-s TR with respect to the first data acquisition time point $\boldsymbol{t_0^{BOLD}}$, that is, 1 s. Moreover, a spatial low-pass filter is applied to all the BOLD images (i.e., IFG, IPL, and SPL/IPS) with full width at half maximum (FWHM) of 4×4 pixels (or mm2) (Figure 8-

Middle left panel). Furthermore, to examine a specific voxel activation, the upper 70 % of the entire voxel that is transformed into the z-space is included; in other words, the lower 30 % of the voxels is considered as no-responsive voxels. It is performed to eliminate the effects of the white noise added in generating the BOLD fMRI signals with the SNR of 65.

3.4 Task conditions

In the current study, a simple geometrical model of the right upper limb having 2 degrees of freedom is used to perform horizontal reaching to grasp task in a two-dimensional plane (Table 1).

Table 1. Anthropometric data and functional range of motion

Dimension name	Demonstrator	Imitator
Right upper arm length	0.33 m	0.16 m
Right forearm length	0.27 m	0.12 m
Shoulder horizontal adduction ($\boldsymbol{\theta_1}$)†	0° to 120°	0° to 120°
Elbow horizontal flexion ($\boldsymbol{\theta_2}$)†	0° to 120°	0° to 120°
Viewpoint‡	-180° to 180°	0° to 180°

†The 0° start position for establishing the degrees of each motion is 90° shoulder abduction and 90° elbow extension, respectively.

‡The viewpoint angle is measured from the Cartesian positive X-axis so that the positive and the negative Y-axis have +90° and -90°, respectively.

The task is performed one by one under 9 conditions that are described by a combination of three $\boldsymbol{y}$-directional distances (i.e., near, middle, and far) and three $\boldsymbol{x}$-directional

position (i.e., left, center, and right), which represents the demonstrator's relative position with respect to the position and viewpoint of the imitator (Figure 9A).

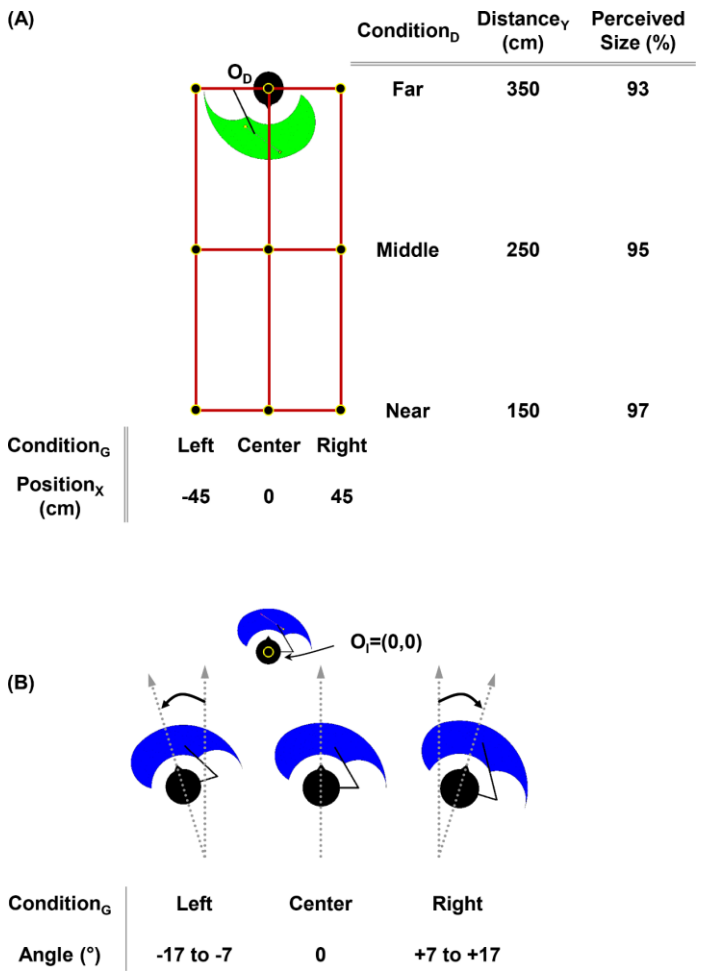


Figure 9. Task condition for a reaching and grasping task. (A) The demonstrator performs the action in each spot represented by a combination of three y-directional (i.e., near, middle, and far) and three x-directional distances (i.e., left, center, and right). The demonstrator always faces toward the negative y-direction, and the imitator faces toward the demonstrator in the origin (0,0). For depth perception, the imitator perceives the demonstrator according to the given perceived size. (B) The imitator turns its body toward the demonstrator by the given angle from 7 to 17 degrees with respect to the positive y-direction.

In particular, each $\boldsymbol{y}$-directional distance is defined as 150 (near), 250 (middle), and 350 cm (far), and each $\boldsymbol{x}$-directional position is set as -45 (left), 0 (center), and 45 cm (right), respectively. Moreover, for more realistic test condition, depth perception is implemented for the imitator to perceive a smaller size of the demonstrator when the demonstrator is further away; specifically, the imitator's depth perception ratios are 97, 95, and 93% of the demonstrator's original body size in near, middle, and far conditions, respectively. In all these conditions, it is assumed that the demonstrator performs the reaching to grasp action while setting its face toward -$\boldsymbol{y}$ direction, whereas the imitator turns its body toward the demonstrator. In this case, the body turn angles are ±7, ±10, and ±17 degrees when the demonstrator is respectively in far, middle, and near position (Figure 9B). For a given condition, two tests are performed; one is an action observation, where the imitator simply observes the demonstrator's action, and the other is an action execution, in which the imitator actually performs the observed action. Finally, the trajectories of each action is generated by a vector-integration-to-endpoint model (Bullock, Bongers, Lankhorst, & Beek, 1999; Bullock & Grossberg, 1988), the duration of which is 1.5 s, and sampled at 128 Hz.

3.5 Assessment criteria for the model performance

After training of all three neural networks, the performance of the proposed MNS model is assessed with respect to the following three criteria: i) learning curves produced by each network, synaptic weights, and the associated magnitude of errors, ii) the kinematic

outputs of the end-effector and its joints, and lastly iii) the response patterns in synthetic BOLD fMRI signals.

3.5.1 Learning curves and associated errors in the neural networks

First of all, for a given condition, the performance of the neural networks modeling functional roles of the SPL/IPS, IFG, and IPL is assessed by means of learning curves produced by the neural networks. Specifically, according to the learning algorithm described in the section 3.2.2, two MSE functions (i.e., from both OLS and GCV methods) are produced by the neural networks for each condition. In particular, it is expected that, when the networks are trained well enough, the MSE of OLS eventually becomes smaller than the MSE of GCV for more than at least two consecutive iterations.

Next, the network weights between the hidden-to-output units are assessed. The network weights represent the local mean synaptic activity of a group of neurons in each brain region. Therefore, the positive and negative weights correspond to the excitatory and inhibitory synaptic activity, respectively.

In addition, the networks are assessed in terms of their prediction quality. Particularly, the root-mean-square (RMS) of the predicted outputs with respect to the expected outputs is employed to measure the error. Therefore, the root-mean-square error (RMSE) is computed for the entire workspace of the demonstrator, which is observed and transformed by the imitator, and in turn, compared with its own workspace in the egocentric frame of reference. It is predicted that the RMSE surface is bounded by a

certain error threshold value (e.g., 0.1 mm), which is considered to be small enough to produce accurate imitation through the SPL/IPS and the MNS.

3.5.2 Behavioral measures (kinematics and mental transformation) of the neural networks

The model performance is also assessed in terms of the quality of imitated action by measuring the horizontal planar movements of the right arm composed of two joints (i.e., shoulder and elbow). Specifically, for kinematic analysis, the spatial trajectory and velocity of the end-effector (i.e., right hand) are examined. In addition, the angular position and velocity of each joint angle are also measured.

Another approach to assess the performance of the proposed MNS model is through the mental rotation, which is examined by the measuring the relationship between the rotation angle and the response time (Bock & Dalecki, 2015; Dalecki, Hoffmann, & Bock, 2012). Particularly, this concept is expanded to include all the visuospatial transformation capabilities of the current SPL/IPS network; namely, the clockwise and counterclockwise rotation as well as the scaling and the translation, although the last two transformations are not explicitly described in the test. Consequently, the response time is defined as the processing time of the SPL/IPS network to successfully transform the observed action from the allocentric to the imitator's egocentric frame of reference. For this analysis, it is assumed that the imitator can see the demonstrator's action through its body.

3.5.3 Simulated BOLD fMRI responses

Finally, the model is assessed by measuring the simulated BOLD fMRI responses for specific given conditions. The simulated BOLD fMRI signals are compared with actual responses reported in other relevant literatures to validate the activation patterns induced by the observation and execution of actions. Moreover, in the simulated BOLD fMRI signals, both view-independent and view-dependent activation patterns are investigated. Particularly, for the analysis, the distance measures as well as the masking paradigm is used to examine the similar (i.e., view-independent MNS) or different (i.e., view-dependent MNS) activation patterns that are engaged in different conditional stimuli. To examine the similarity of two BOLD images in two-dimensional matrix forms, the covariance matrix distance (CMD; $\boldsymbol{d_{COV}}$) is devised based on the idea of correlation matrix distance ($\boldsymbol{d_{cor}}$) (Herdin, Czink, Ozcelik, & Bonek, 2005). Although $\boldsymbol{d_{cor}}$ is informative, it produces a not-a-number when any input vector (i.e., A or B) is stationary; in other words, when any standard deviation (i.e., $\boldsymbol{\sigma}_{\mathrm{A}}$ or $\boldsymbol{\sigma}_{\mathrm{B}}$) is zero because the correlation matrix of these two vectors $\boldsymbol{cor}$(A,B) is defined as $\boldsymbol{cor}(\mathrm{A,B}) = \boldsymbol{cov}(\mathrm{A,B})/(\boldsymbol{\sigma}_{\mathrm{A}}\boldsymbol{\sigma}_{\mathrm{B}})$. Therefore, $\boldsymbol{d_{COV}}$ is defined as:

$$\boldsymbol{d_{cov}}(\mathrm{X,Y}) = \mathbf{1} - \boldsymbol{tr}(\mathrm{XY})/(\|\mathrm{X}\|_{\boldsymbol{F}}\|\mathrm{Y}\|_{\boldsymbol{F}}) \tag{36}$$

where X and Y are the covariance matrices between two BOLD responses in each condition of a given task, $\boldsymbol{tr}(\cdot)$ is the matrix trace, and $\|\cdot\|_{\boldsymbol{F}}$ is the Frobenius norm. The CMD becomes 0 if the matrices are similar (i.e., their eigenvalues are same; see Appendix C), whereas it is 1 if they differ to a maximum extent.

Chapter 4: Results

In this chapter, the simulated MNS model is assessed by examining the performance of the computational processes in the SPL/IPS, IFG, and IPL, which are implemented with RBF networks and the learning algorithms based on the OLS and GCV. Specifically, the assessment is conducted at the network, behavioral, and neurophysiological levels according to the criteria for the model performance as described in the section 3.5. Overall, the simulation results based on both batch and online learning methods reveal that each of the computational components modeled the SPL/IPS, IFG and IPL can successfully learn the visuospatial transformation, inverse model, and the forward model, respectively. In particular, after the networks are trained, the results show that i) the imitator can successfully imitate the arm kinematics performed by the demonstrator in a similar way to those observed in humans; ii) the visuospatial transformation allowing for observing the actions in a egocentrically transformed manner leads to functionally similar patterns to those observed during mental transformation of human body as well as mental rotation of objects under various viewpoints, and iii) in agreement with neurophysiological studies, the simulated neural activities during both action observation and imitation are comparable, and more importantly, reveal two types of neural populations encoding view-independent and view-dependent representations of the observed action, respectively. It is critical to note that for the sake of clarity, only three key conditions (i.e., left-middle, center-middle, and right-middle) are depicted in the

results below, however very similar findings are obtained for the other six conditions both in terms of the prescribed three assessment criteria (see the section 3.5).

4.1 Learning of transformation primitives

As illustrated in Figure 7, the four transformation primitives (i.e., clockwise and counterclockwise rotations as well as translation and scaling transformations) are first trained with the batch update method.

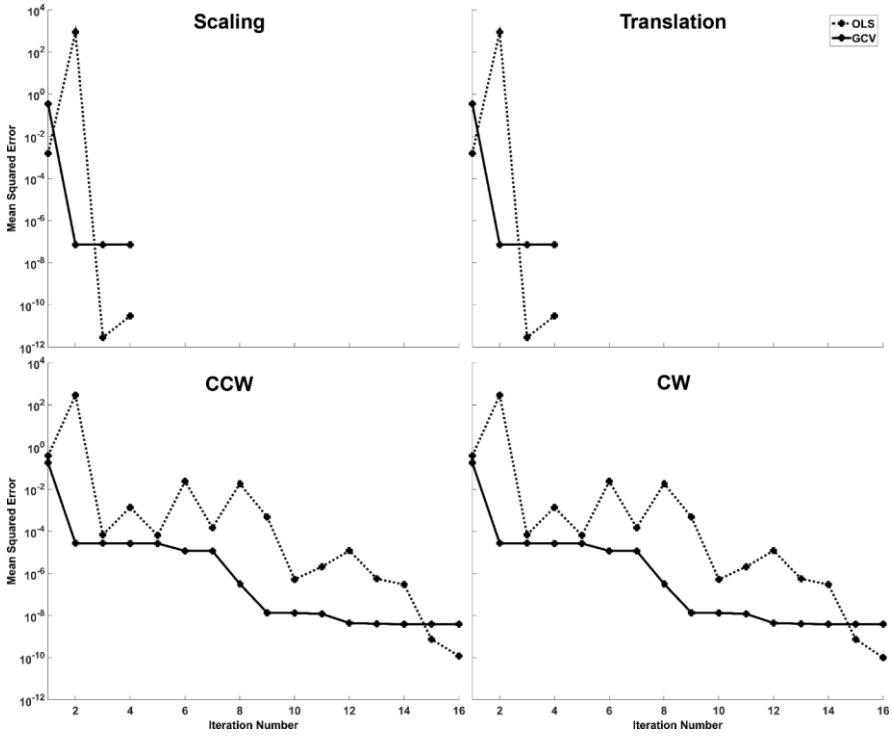


Figure 10. Learning curves of the four transformation primitives in each dimension of the two-dimensional Cartesian coordinates. The scaling and translation networks have similar learning curve slopes. Similarly, the counterclockwise (CCW) and clockwise (CW) rotation networks are similar in their slopes. The dashed lines indicate the MSE in the OLS, whereas the solid lines correspond to the MSE in the GCV.

After training, the results reveal that all the learning curves (i.e., MSEs in both OLS and GCV) commonly converge on a very small number ($\leq \mathbf{7.28 \times 10^{-8}}$; Table 2; Figure 10) with up to 16 RBFs per each dimension of the two-dimensional Cartesian coordinates (Figure 10). Interestingly, the MSE curves for the scaling and translation transformations are very similar to each other in both OLS and GCV, and those curves for the counterclockwise (CCW) and clockwise (CW) rotations are also very similar to each other (Figure 10). Specifically, the scaling and translation networks achieve the similar MSEs (MSE-OLS: $\mathbf{3.00 \times 10^{-11}}$ and MSE-GCV: $\mathbf{7.28 \times 10^{-8}}$) with only 4 RBFs per each dimension, while the CCW (MSE-OLS: $\mathbf{1.22 \times 10^{-10}}$ and MSE-GCV: $\mathbf{3.92 \times 10^{-9}}$) and the CW (MSE-OLS: $\mathbf{1.02 \times 10^{-10}}$ and MSE-GCV: $\mathbf{3.92 \times 10^{-9}}$) networks also reach similar MSEs with 16 RBFs per each dimension, respectively (Table 2). Across the conditions, the standard deviations of both MSE-OLS and MSE-GCV are commonly less than or equal to $\mathbf{1.44 \times 10^{-14}}$. Moreover, for each transformation network, the RBFs with optimal radii are evenly distributed in $[\mathbf{-1}, \mathbf{1}] \times [\mathbf{-1}, \mathbf{1}]$ as explained in the section 3.2.1. For example, for each scaling and translation network, the selected 4 RBFs are placed on each point of $\{(\mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{-1}), (\mathbf{0}, \mathbf{1}), (\mathbf{0}, \mathbf{-1})\}$ and their radii are 2. Similarly, each of the clockwise and counterclockwise rotations is represented with evenly distributed 16 RBFs with the radius of 2.

Table 2. Mean last MSEs of the transformation primitives in the batch update method

MSE	Scaling	Translation	CCW	CW
MSE-OLS	$\mathbf{3.00 \times 10^{-11}}$	$\mathbf{3.00 \times 10^{-11}}$	$\mathbf{1.22 \times 10^{-10}}$	$\mathbf{1.02 \times 10^{-10}}$
MSE-GCV	$\mathbf{7.28 \times 10^{-8}}$	$\mathbf{7.28 \times 10^{-8}}$	$\mathbf{3.92 \times 10^{-9}}$	$\mathbf{3.92 \times 10^{-9}}$

Moreover, since the transformation primitives are independent of the relative spatial relationships between the demonstrator and the imitator, those four primitive networks should produce very similar output results regardless of the testing conditions. Specifically, it could be demonstrated by the average RMSE values for the aforementioned nine conditions: the scaling ($\mathbf{9.10 \times 10^{-6}}$ m), translation ($\mathbf{9.10 \times 10^{-6}}$ m), CCW ($\mathbf{1.40 \times 10^{-6}}$ m), and CW ($\mathbf{1.43 \times 10^{-6}}$ m), respectively (Figure 11).

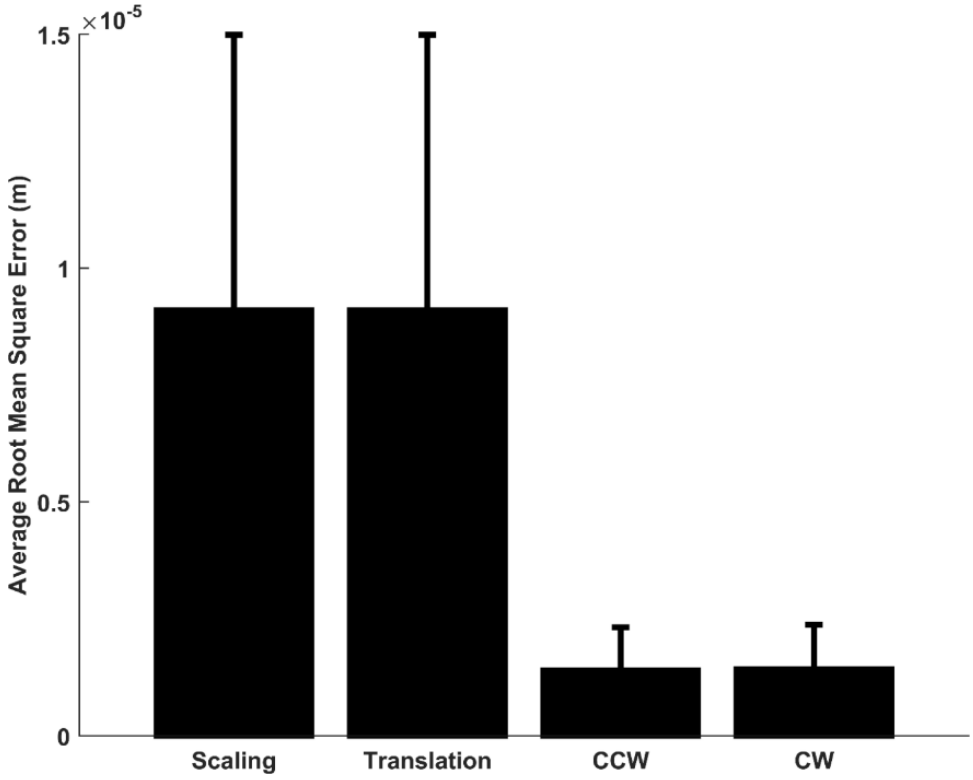


Figure 11 Average root mean square error of four transformation primitives. The average RMSE of these transformation networks is less than 1.0×10-5 m or 0.01 mm.

As a result, the average RMSE is less than $\mathbf{1.0 \times 10^{-5}}$ m (or 0.01 mm) in all transformation primitives (Figure 11). Therefore, these four accurate transformation primitive networks can result in good performance for the SPL/IPS network (i.e., a combination network of these four subnetworks) after the learning of the SPL/IPS is completed.

4.2 Learning curves and associated errors in the neural networks

4.2.1 Batch and online update methods

As stated in section 3.2.4, the performances of the proposed MNS model that is based on both batch and online update methods are examined under same conditions. It must be noted that although the MSE curves in both OLS and GCV are actually used for the training of the neural networks as explained in section 3.2.2, the average of these two MSE curves having overall similar trends is depicted for the sake of simplicity (Figure 12).

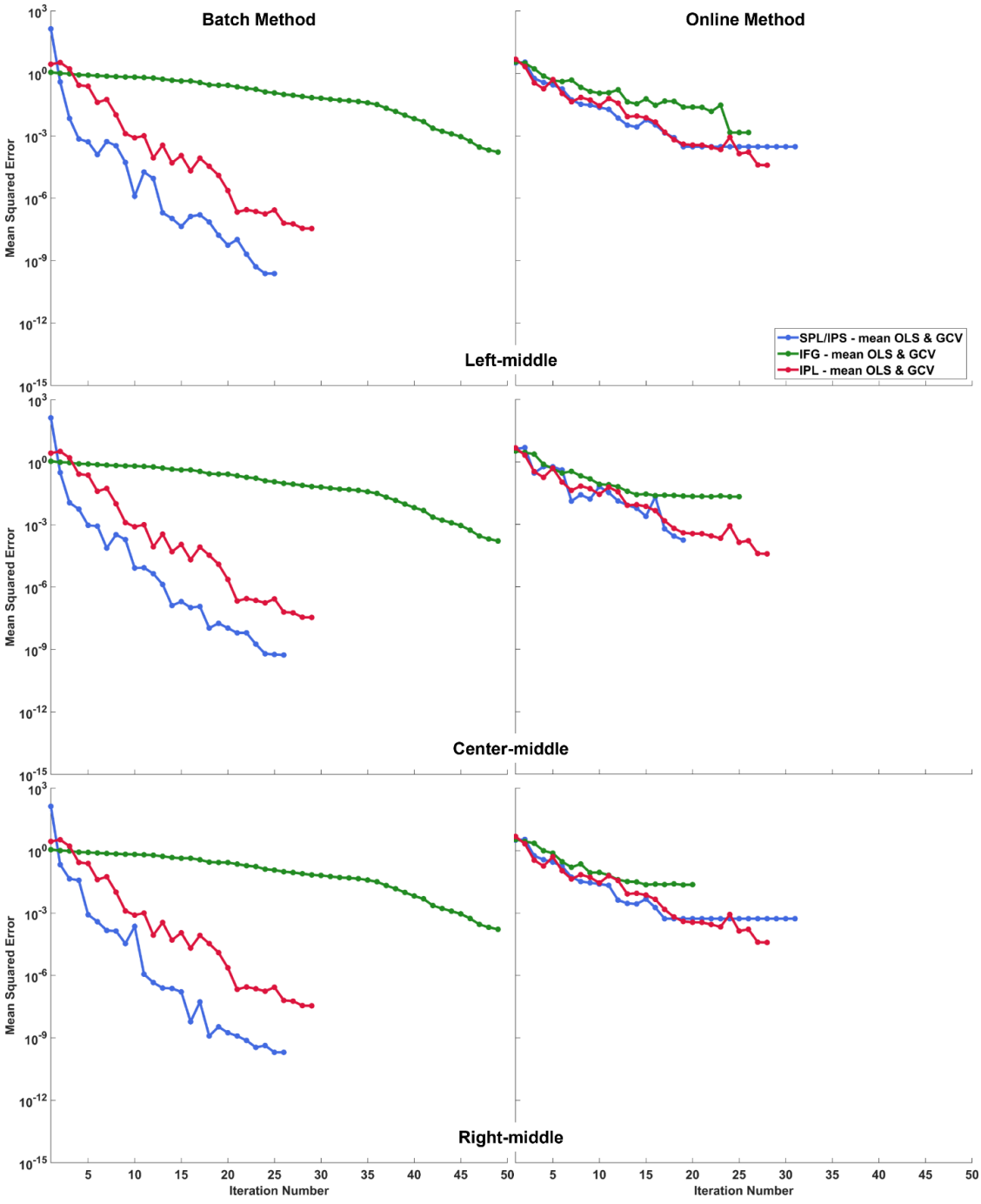


Figure 12. Mean learning curves of the three neural networks using the batch (left column) and online (right column) update methods. The demonstrator performs an action in the left-middle (first row), center-middle (second row), and right-middle (third row) conditions. The blue, green, and red lines represent the mean learning curves for the SPL/IPS, IFG, and IPL, respectively.

First, the batch update method results in very similar patterns particularly for the SPL/IPS and the IPL across conditions, albeit some minor differences for the IFG (Figure 12-Left column). Although the details of the learning curves are different across conditions, the overall results reveal that i) all three neural networks are successfully trained with pretty low MSE values, ii) the SPL/IPS tends to learn faster than the IPL, and the IPL tends to learn faster than the IFG, and iii) the IFG requires a larger number of RBFs (49 RBFs per dimension) to model the inverse model compared to the IPL (29 RBFs per dimension) for the forward model and the SPL/IPS (26 RBFs per dimension) for the visuospatial transformation. Particularly, after training is completed, the last MSE values (i.e., MSE-OLS and MSE-GCV) of the SPL/IPS are slightly different across conditions, whereas those of the IFG and IPL are same (Table 3). This result reveals that all neural networks commonly converge on a small MSE ($\leq \mathbf{3.83 \times 10^{-4}}$) regardless of the testing conditions and the MSE types.

Table 3. Last MSEs of the neural networks that are trained with the batch update method

Network	MSE	Left-middle	Center-middle	Right-middle
SPL/IPS	MSE-OLS	$\mathbf{1.75 \times 10^{-12}}$	$\mathbf{2.71 \times 10^{-13}}$	$\mathbf{4.31 \times 10^{-14}}$
	MSE-GCV	$\mathbf{4.73 \times 10^{-10}}$	$\mathbf{6.37 \times 10^{-11}}$	$\mathbf{8.03 \times 10^{-11}}$
IFG	MSE-OLS	$\mathbf{1.24 \times 10^{-6}}$	$\mathbf{1.24 \times 10^{-6}}$	$\mathbf{1.24 \times 10^{-6}}$
	MSE-GCV	$\mathbf{3.83 \times 10^{-4}}$	$\mathbf{3.83 \times 10^{-4}}$	$\mathbf{3.83 \times 10^{-4}}$
IPL	MSE-OLS	$\mathbf{2.39 \times 10^{-10}}$	$\mathbf{2.39 \times 10^{-10}}$	$\mathbf{2.39 \times 10^{-10}}$
	MSE-GCV	$\mathbf{1.26 \times 10^{-8}}$	$\mathbf{1.26 \times 10^{-8}}$	$\mathbf{1.26 \times 10^{-8}}$

Although less pronounced than for the batch method, the online update method results in a similar trend regarding the speed of the SPL/IPS, IPL, and IFG (Figure 12-Right column). Namely, the SPL/IPS tends to be slightly faster than the IPL, and the IPL tends

to be must faster than the IFG independently of the testing conditions. Moreover, although the IPL reveals a similar learning slope across conditions, the learning curves of the SPL/IPS and IFG are different. Furthermore, although the MSEs are remarkably larger compared to the batch update method (Table 3 and 4), the MSEs in the online update method are still small ($\leq \mathbf{1.63 \times 10^{-2}}$) across conditions (Table 4). This is not surprising considering the learning procedure, in which all three neural networks are interacting each other, thus the accuracy of a certain network is also affected by the errors from the other two networks.

Table 4. Last MSEs of the neural networks that are trained with the online update method

Network	MSE	Left-middle	Center-middle	Right-middle
SPL/IPS	MSE-OLS	$\mathbf{6.15 \times 10^{-6}}$	$\mathbf{3.57 \times 10^{-5}}$	$\mathbf{6.15 \times 10^{-6}}$
	MSE-GCV	$\mathbf{1.46 \times 10^{-5}}$	$\mathbf{3.18 \times 10^{-4}}$	$\mathbf{1.46 \times 10^{-5}}$
IFG	MSE-OLS	$\mathbf{1.06 \times 10^{-2}}$	$\mathbf{1.30 \times 10^{-3}}$	$\mathbf{1.80 \times 10^{-3}}$
	MSE-GCV	$\mathbf{1.63 \times 10^{-2}}$	$\mathbf{1.40 \times 10^{-3}}$	$\mathbf{1.60 \times 10^{-3}}$
IPL	MSE-OLS	$\mathbf{3.27 \times 10^{-6}}$	$\mathbf{3.27 \times 10^{-6}}$	$\mathbf{3.27 \times 10^{-6}}$
	MSE-GCV	$\mathbf{8.32 \times 10^{-6}}$	$\mathbf{8.32 \times 10^{-6}}$	$\mathbf{8.32 \times 10^{-6}}$

In addition to the comparison of the learning curves between the batch and online update methods, the RMSE surfaces and their mean values (including standard deviations) are compared for a more direct assessment (Figure 13).

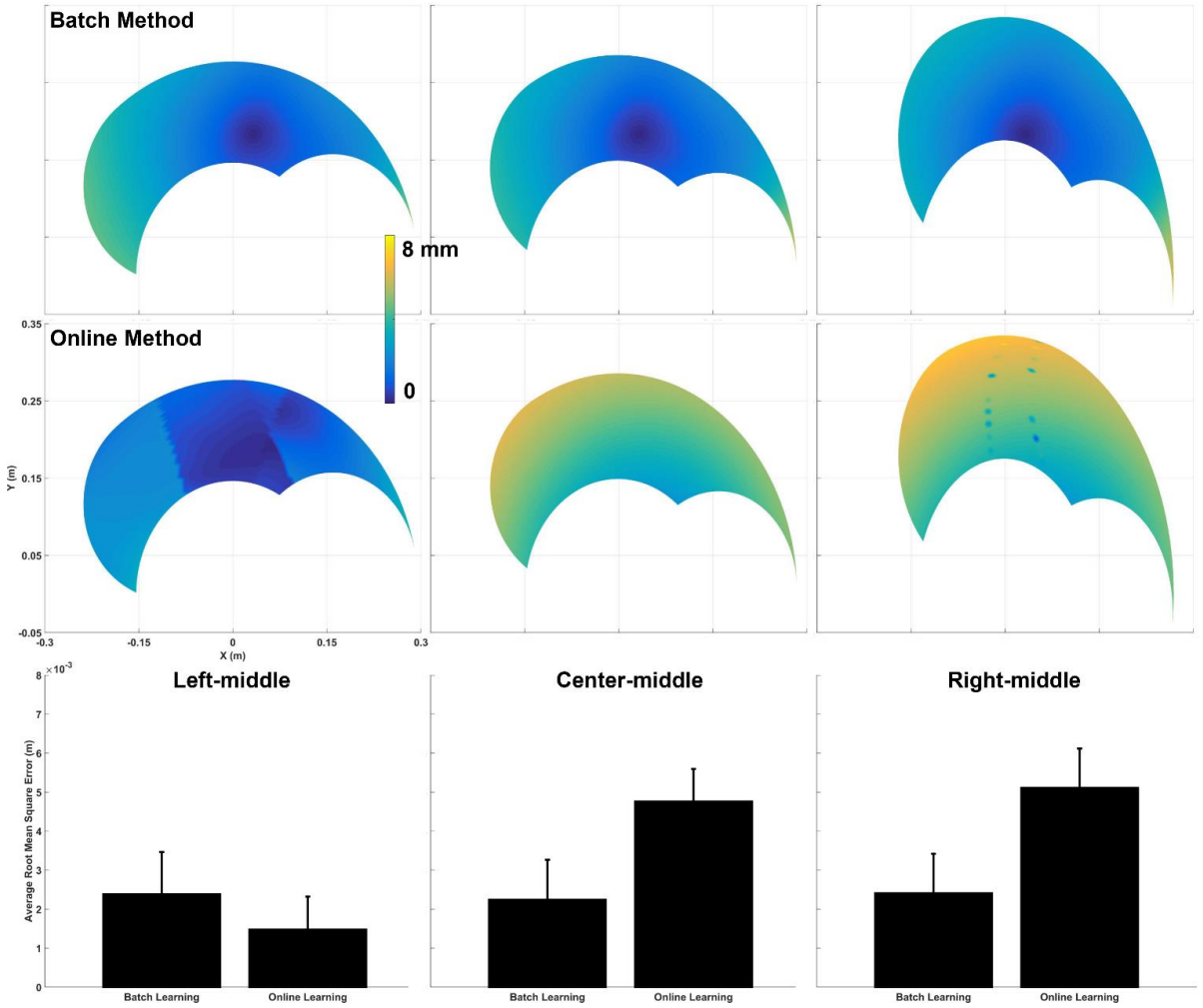


Figure 13. RMSE surfaces and mean RMSE in batch (first row) and online (second row) update methods under the left-middle (first column), center-middle (second column), and right-middle (third column) conditions. The RMSE surfaces are ranged from 0 (blue) to 8 mm (yellow). The mean RMSE (third row) is depicted with the mean and standard deviation.

For both batch and online methods, the RMSE surfaces have minimum values near the lower center (or origin) point where the imitator is placed, and maximum values around the outermost corner where the imitator can reach by stretching its arm (Figure 13, first two rows). More importantly, the mean RMSE values are similar between the batch and online update methods (Figure 13, third row). Particularly, it shows that the mean RMSE values are less than 8 mm regardless of the testing conditions. Moreover, in the left-middle condition, even online update method results in lower mean RMSE value than the batch update method (Figure 13; Table 5). As a result, this demonstrates that the accuracy of the online update method is comparable to the one obtained with the batch method (Table 5) while the former allows three neural networks to interact during learning.

Table 5. Mean RMSE in the batch and online update methods (mean ± standard deviation)

Method	Left-Middle	Center-Middle	Right-Middle
Batch	$(\mathbf{2.40 \pm 1.10}) \times \mathbf{10^{-3}}$	$(\mathbf{2.20 \pm 1.10}) \times \mathbf{10^{-3}}$	$(\mathbf{2.40 \pm 1.10}) \times \mathbf{10^{-3}}$
Online	$(\mathbf{1.50 \pm 0.80}) \times \mathbf{10^{-3}}$	$(\mathbf{4.80 \pm 0.80}) \times \mathbf{10^{-3}}$	$(\mathbf{5.10 \pm 1.00}) \times \mathbf{10^{-3}}$

The mean RMSE of the online method is about twice its counterpart of the batch method (Table 5), but it is still small because the mean RMSE is less than or equal to 5.10 mm regardless of the testing conditions (Table 5). Moreover, the standard deviations of the mean RMSEs are very similar to each other in both batch and online methods across conditions.

4.2.2 Network weights

In this section, the neural weights of the RBF networks are examined for both action observation and execution.

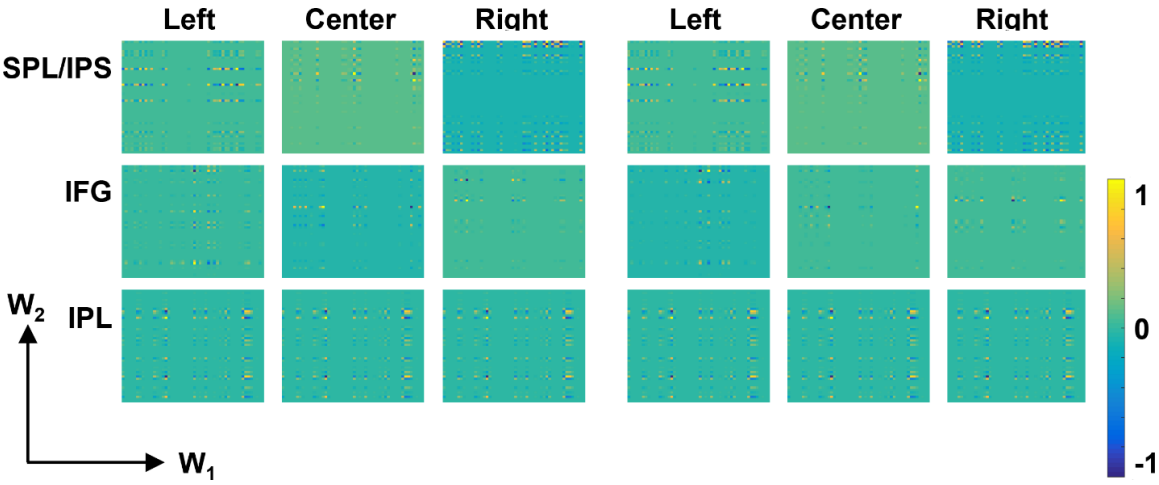


Figure 14. Neural weights of all three networks in action observation (left panel) and execution (right panel) under the left-middle, center-middle, and right-middle conditions. For visualization purpose, the weights are depicted in two dimensional space of 50×50 RBFs, each of which can represent X and Y dimension in the Cartesian coordinates. The weights are normalized to be scaled from -1 (blue) to 1 (yellow).

As expected, the neural weights are similar for both observation and execution regardless of the testing conditions (Figure 14). Specifically, the results reveal that the activity patterns of both excitatory and inhibitory synapses are very similar for both action observation and execution (Figure 14). Moreover, the synaptic patterns of the IPL are same independently of the testing conditions, whereas the SPL/IPS and IFG have slightly different synaptic patterns in each condition.

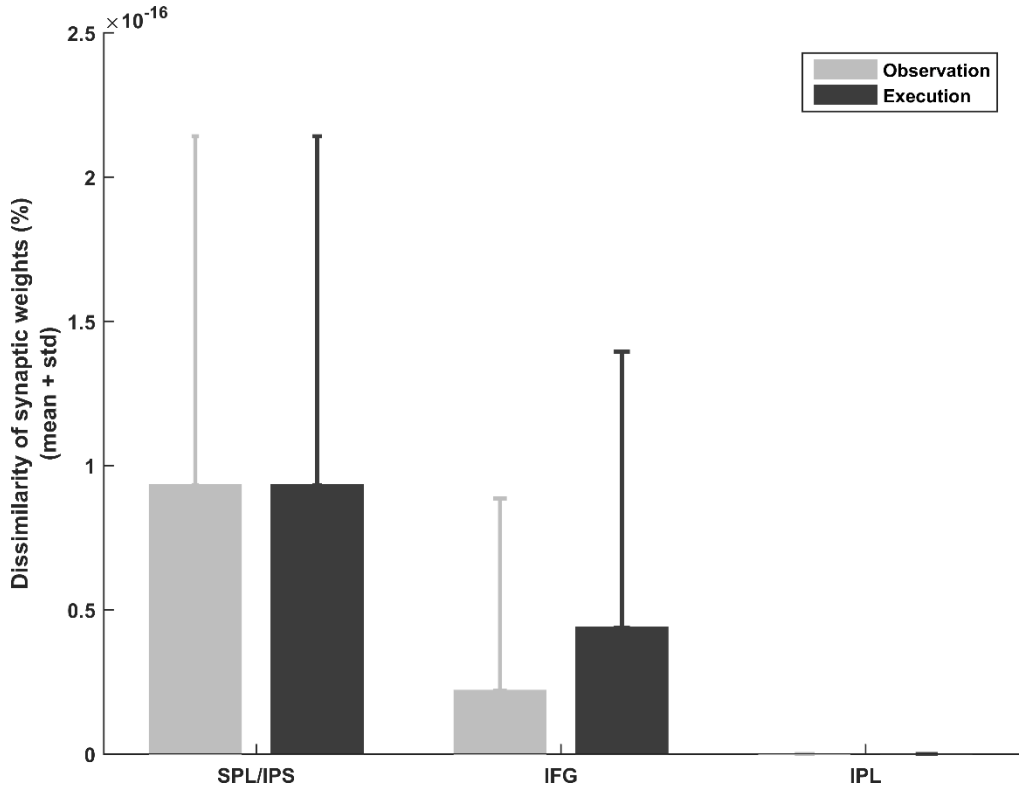


Figure 15. Mean dissimilarity of the synaptic weights of each region in action observation and execution. The dissimilarity measure based on the covariance matrix distance (CMD) quantifies the dissimilarity between 0 (identical) and 1 (different).

Furthermore, the dissimilarity of the mean synaptic activities between action observation and execution is quantitatively observed by the CMD. The results reveal that the average synaptic weights of each neural network are almost same across conditions during action observation and execution (Figure 15). Although the IFG reveals a slight difference in the mean synaptic weights, such a discrepancy is still very small ($\leq \mathbf{2.19 \times 10^{-17}}$). As a result, all the results indicate that all three brain regions would similarly activate during action observation and execution.

4.2.3 RMSE of the network outputs

For action observation and execution, the RMSE surfaces and their mean values (including standard deviations) are also obtained in the left-middle, center-middle, and right-middle conditions (Figure 16).

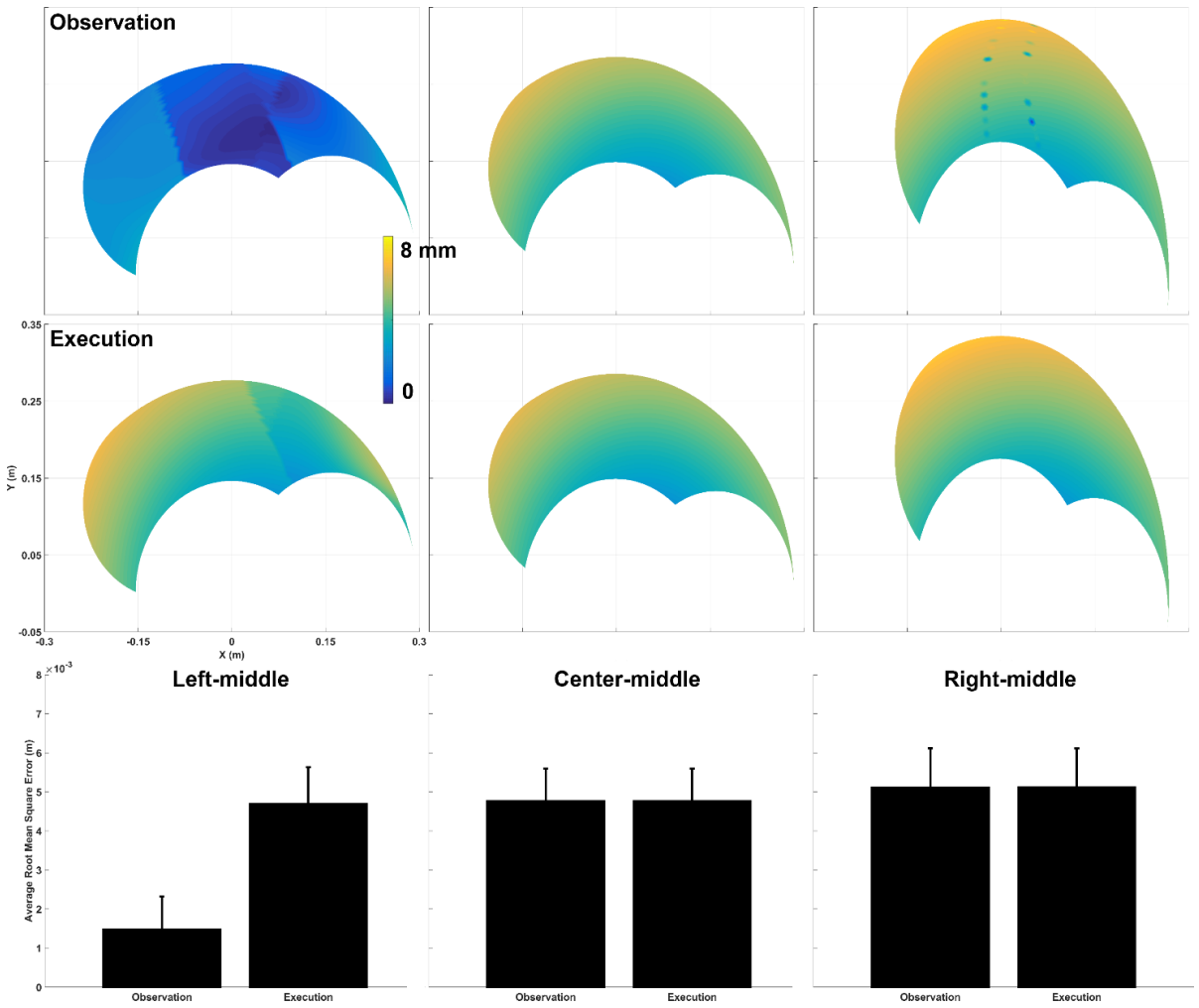


Figure 16. RMSE surfaces and mean RMSE in action observation (first row) and execution (second row) under the left-middle (first column), center-middle (second column), and right-middle (third column) conditions. The RMSE surfaces are ranged from 0 (blue) to 8 mm (yellow). The mean RMSE (third row) is depicted with the mean and standard deviation.

These RMSE surfaces also have minimum values near the lower center (or origin) point where the imitator is placed, and maximum values around the outermost corner where the imitator can reach by stretching its arm (Figure 16, first two rows). Although the RMSE surfaces are different between action observation and execution in the left-middle condition, the discrepancy is still very small ($\leq \mathbf{3.20}$ mm; Table 6). Therefore, it reveals that the RMSE surfaces are very similar in action observation and execution regardless of the testing condition (Figure 16).

Table 6. Mean RMSE in action observation and execution (mean ± standard deviation)

Condition	Left-Middle	Center-Middle	Right-Middle
Observation	$(\mathbf{1.50 \pm 0.85}) \times \mathbf{10^{-3}}$	$(\mathbf{4.80 \pm 0.83}) \times \mathbf{10^{-3}}$	$(\mathbf{5.10 \pm 1.00}) \times \mathbf{10^{-3}}$
Execution	$(\mathbf{4.70 \pm 0.94}) \times \mathbf{10^{-3}}$	$(\mathbf{4.80 \pm 0.83}) \times \mathbf{10^{-3}}$	$(\mathbf{5.10 \pm 1.00}) \times \mathbf{10^{-3}}$

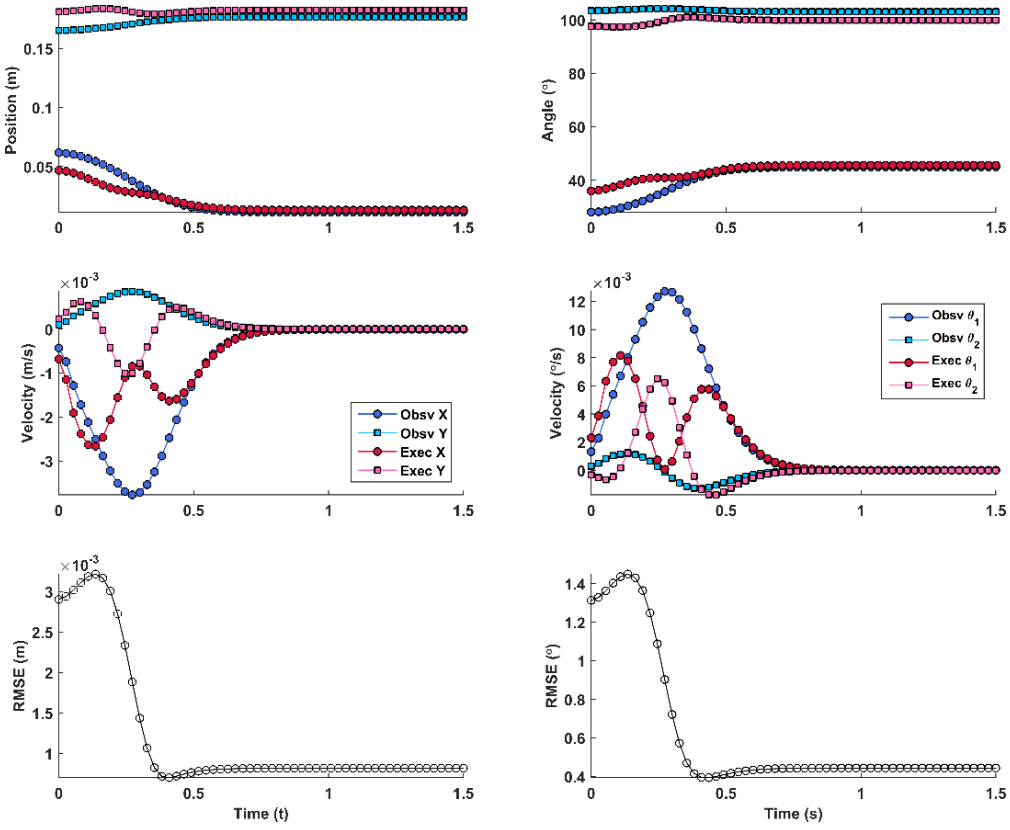
The mean RMSE values are less than or equal to 5.10 mm regardless of the testing conditions, and the corresponding standard deviations are also very small ($\leq \mathbf{1.00} \times \mathbf{10^{-3}}$; Table 6).

4.3 Behavioral measures of the neural networks

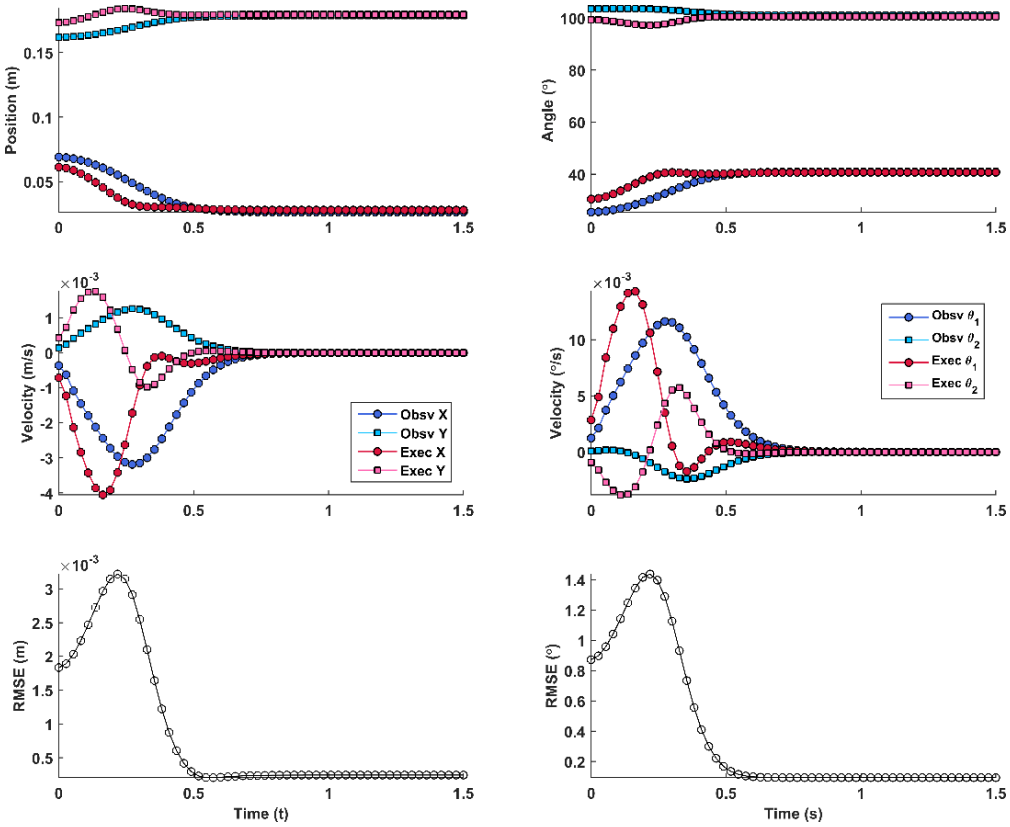
4.3.1 Kinematics of the end-effector and joints

The inverse and forward kinematics respectively in the IFG and IPL are investigated in this section during action observation and execution in the left-middle, center-middle, and right-middle conditions.

A)



B)


Position (m)
Angle (°)
Velocity (m/s)
Velocity (°/s)
Obsv X
Obsv Y
Exec X
Exec Y
Obsv θ_1
Obsv θ_2
Exec θ_1
Exec θ_2
RMSE (m)
RMSE (°)
Time (t)
Time (s)

C)

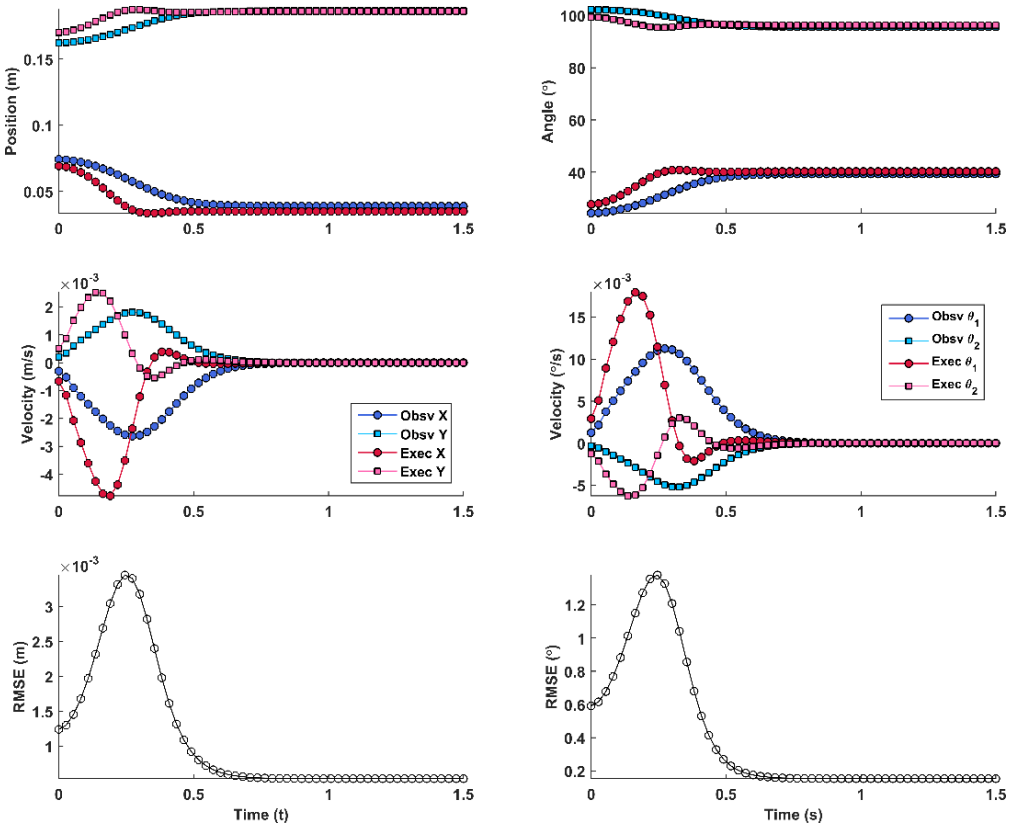

Position (m)
Angle (°)
Velocity (m/s)
Velocity (°/s)
Obsv X
Obsv Y
Exec X
Exec Y
Obsv θ_1
Obsv θ_2
Exec θ_1
Exec θ_2
RMSE (m)
RMSE (°)
Time (t)
Time (s)

Figure 17. Kinematics of the end-effector and its joint angles under (A) the left-middle, (B) center-middle, and (C) right-middle conditions. In each figure, the left and right panels depict positions vs. time and joint angles vs. time, respectively. Each row corresponds to the position (or angle), its first derivative (i.e., velocity), and its mean RMSE curves, respectively. The dark blue (X-axis) and light blue (Y-axis) lines are for the action observation, whereas the dark red (shoulder horizontal adduction) and light red (elbow horizontal flexion) lines are for the action execution.

Overall, the observed and imitated trajectories are similar while the simulated kinematics generated by the model are similar to those observed in humans (i.e., sigmoid-shape displacement as well as bell-shaped and single-peaked velocity profiles). It must be noted that as the action is executed the imitated action becomes closer to the observed action. This suggests that the imitator will correct its trajectory error during action as shown in the RMSE curves, thus resulting in the presence of a second peak in the velocity profiles (Figure 17).

Table 7. Mean RMSE of the stabilized kinematics between the observed and executed actions (mean ± standard deviation)

Kinematics	Left-Middle	Center-Middle	Right-Middle
Position (m)	$\mathbf{8.15 \times 10^{-4}}$ $\mathbf{\pm 1.76 \times 10^{-8}}$	$\mathbf{2.43 \times 10^{-4}}$ $\mathbf{\pm 6.31 \times 10^{-8}}$	$\mathbf{5.35 \times 10^{-4}}$ $\mathbf{\pm 1.29 \times 10^{-7}}$
Joint angle (°)	$\mathbf{4.43 \times 10^{-1}}$ $\mathbf{\pm 1.60 \times 10^{-7}}$	$\mathbf{0.92 \times 10^{-1}}$ $\mathbf{\pm 1.52 \times 10^{-8}}$	$\mathbf{1.53 \times 10^{-1}}$ $\mathbf{\pm 2.98 \times 10^{-7}}$

As shown in Figure 17, the imitated action becomes stabilized after 1 second regardless of the testing condition. Therefore, the mean RMSE of the kinematics between the observed and executed actions are quantified from 1.0 to 1.5 seconds (Table 7). The results reveal that the imitator can accurately reproduce the observed action with very small mean RMSE values in terms of the position of the end-effector ($\leq \mathbf{0.815}$ mm) and its joint angles ($\leq \mathbf{0.443}°$) independently of the testing conditions.

4.3.2 Mental transformation

The performance of the SPL/IPS is measured by considering the response time during mental rotation as stated in 3.5.2.

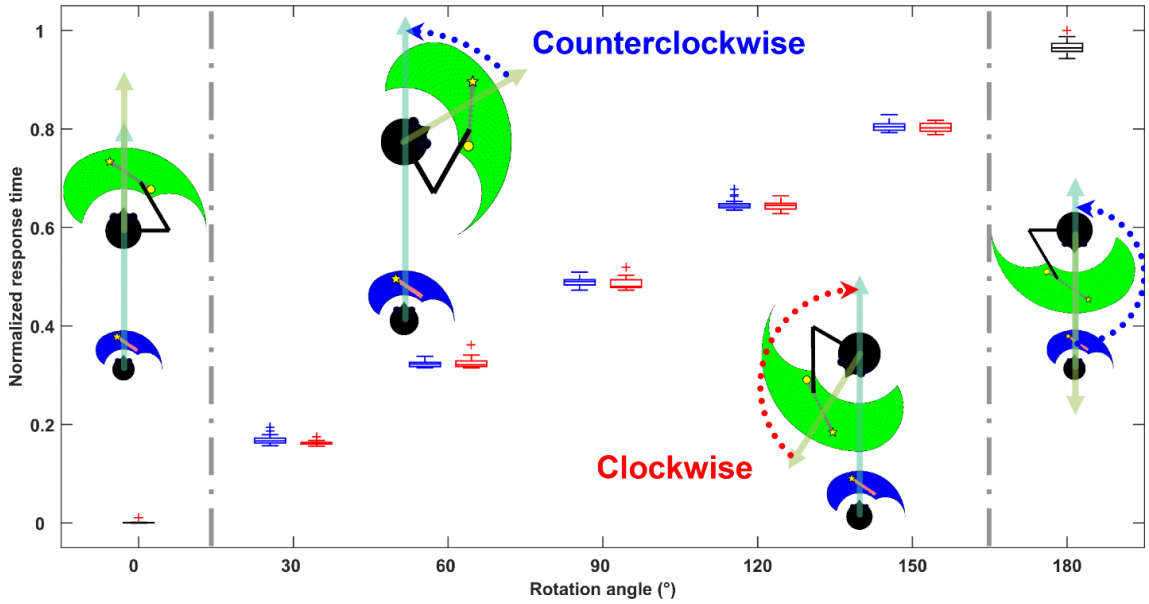


Figure 18. Normalized response time during mental rotation. The mental rotation is performed in clockwise (red) and counterclockwise (blue) directions for a given angle between 30 to 150 degrees. The mean and the standard deviation are calculated on 20 trials for each case. The outliers are marked with a plus sign (+).

The findings reveal that the normalized response time required to transform the observed action monotonically increases with respect to the amount of rotation angle. It must be noted that this transformation includes rotation, translation, and scaling, therefore, the response time actually reflects the processing time in the SPL/IPS. Therefore, for the special case of 0 degree, the SPL/IPS performs only the translation and scaling, and for the other special case of 180 degree, the SPL/IPS employs the counterclockwise subnetwork for the mental rotation (Figure 18).

4.4 Simulated BOLD fMRI responses

4.4.1 BOLD fMRI responses during action observation and execution

While the previous sections focus on the validation of the proposed neural model at the neural network and behavioral levels, this section focuses on neurophysiological assessments by employing the synthetic BOLD fMRI responses generated by the three main neural networks (i.e., SPL/IPS, IFG, and IPL).

As expected, the results reveal that during action observation and execution the simulated BOLD fMRI responses from three neural networks are similar (Figure 19). More specifically, the SPL/IPS network generates different BOLD pattern across condition, however for a given condition its patterns are very similar for both action observation and execution. On the other hand, the BOLD fMRI patterns produced by the IPL are same

independently of the testing conditions as well as during action observation and execution.

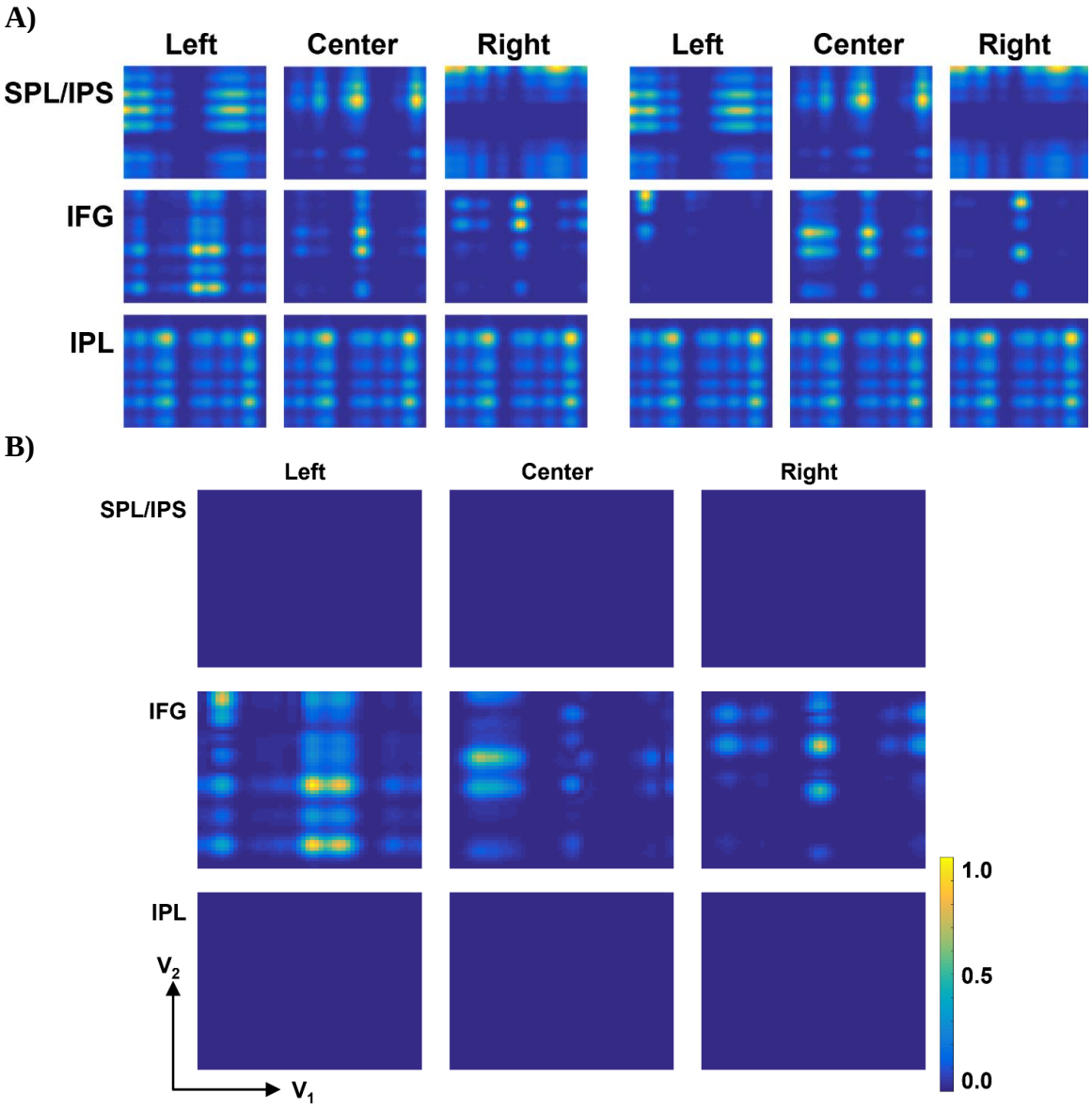


Figure 19. Simulated BOLD fMRI response during action observation and execution. (A) The raw simulated BOLD fMRI responses of three neural networks during action observation (left panels) and execution (right panels) in three middle condition. (B) The difference of BOLD responses during action observation and execution. The BOLD response is scaled between 0 and 1. The horizontal and vertical axes correspond to the first and second dimensional voxels, respectively.

Interestingly, although the synthetic BOLD responses generated by the IFG are relatively different between testing conditions (the details will be discussed in the next section), the overall activities of the virtual voxels reveal fairly similar patterns between action observation and execution, in spite of somewhat different magnitudes. This is confirmed quantitatively with the dissimilarity measurement, which reveals that the mean BOLD fMRI responses simulated during action observation and execution are very similar (Figure 20).

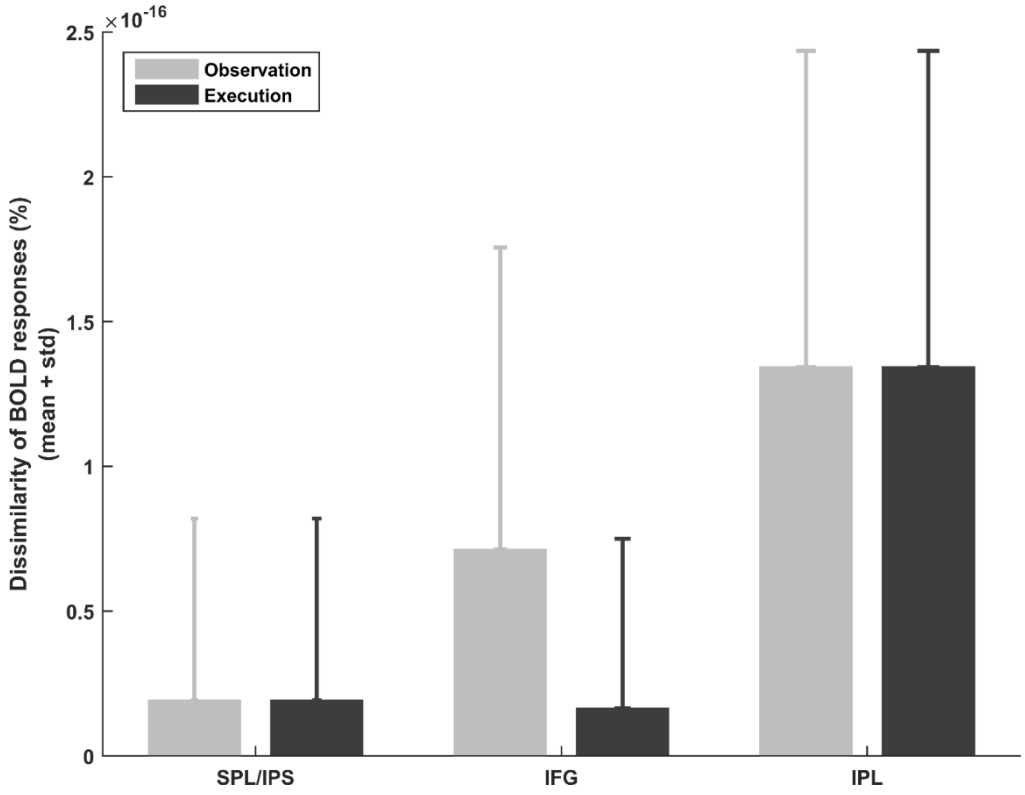


Figure 20. Dissimilarity measures of the simulated BOLD fMRI response during action observation and execution. The covariance matrix distance is employed to measure the dissimilarity. Two BOLD responses are similar when the value is closer to 0, whereas they are different when the value is closer to 1.

4.4.2 View-independent vs. view-dependent MNS activity

One important element in the validation of this work is to assess the emergence of the view-independent and view-dependent MNS activities. When examining the activation of all the voxels under various testing conditions, it appears that some voxels are activated independently of the condition, whereas others are activated only in a specific condition, although their strengths are different (Figure 21).

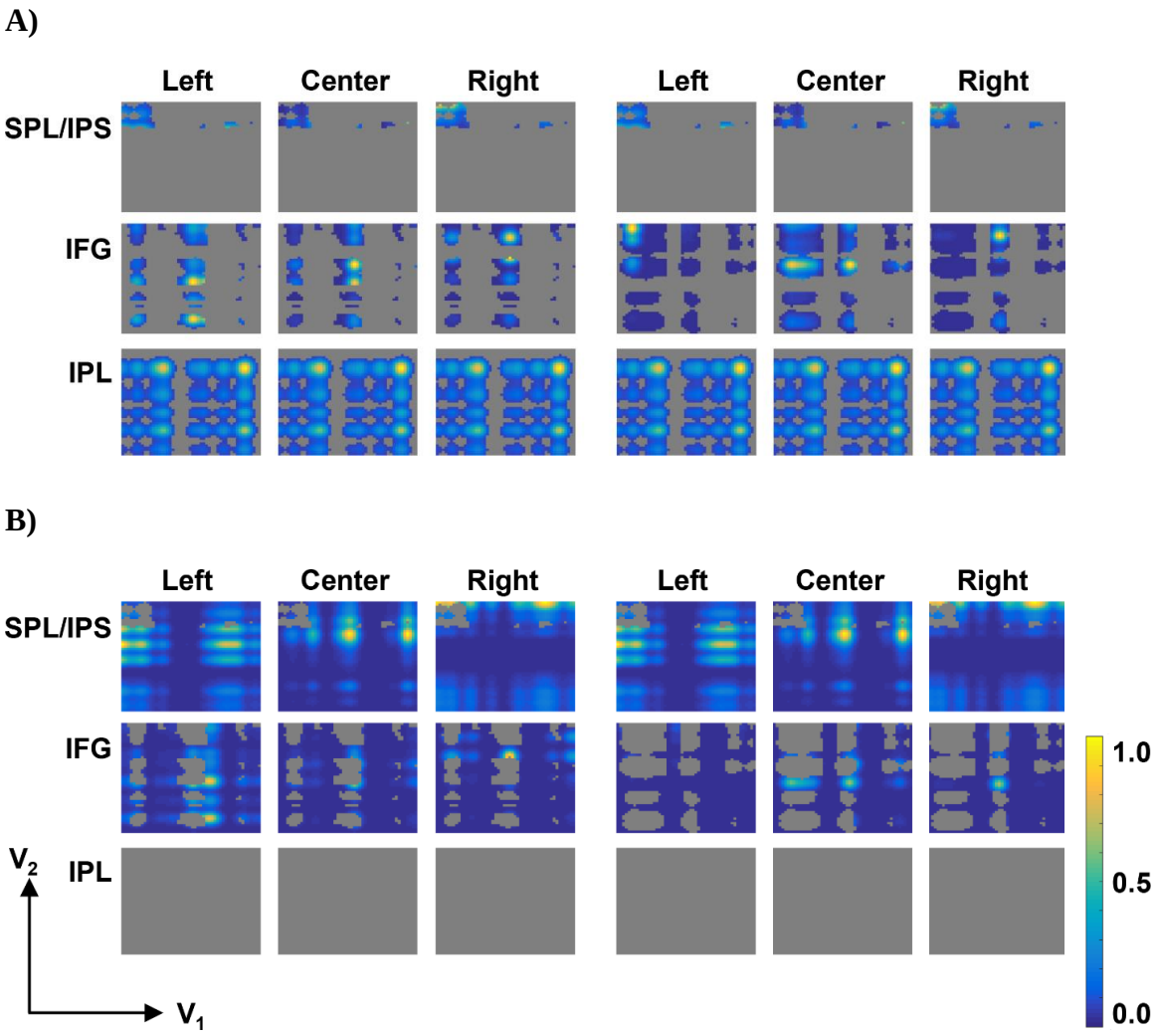


Figure 21. The view-independent and view-dependent BOLD fMRI responses during action observation and execution. (A) The view-independent BOLD fMRI responses during action observation (left panels) and execution (right panels) in three middle conditions. (B) The view-dependent BOLD fMRI responses also during action observation (left panels) and execution (right panels) in three middle conditions. The BOLD responses are scaled between 0 and 1, and the gray color describes a mask to emphasize the corresponding voxels. The horizontal and vertical axes correspond to the first and second dimensional voxels, respectively.

Specifically, the findings reveal that the neural networks modeling the SPL/IPS and IFG are able to generate synthetic fMRI activities that would correspond to view-independent and view-dependent neural populations (see Figure 19A, first two rows). Conversely, the IPL network is able to produce synthetic fMRI activities that correspond only to the view-independent neural population (see Figure 19A, third row).

Table 8. The mean ratio for the view-independent and view-dependent voxels in action observation and execution

Region	Voxel Type	Observation	Execution	Mean
SPL/IPS	View-independent voxels	**4.36%**	**4.36%**	**4.36%**
	View-dependent voxels	**95.64%**	**95.64%**	**95.64%**
IFG	View-independent voxels	**20.16%**	**36.24%**	**28.2%**
	View-dependent voxels	**79.84%**	**63.76%**	**71.8%**
IPL	View-independent voxels	**100.00%**	**100.00%**	**100.00%**
	View-dependent voxels	**0.00%**	**0.00%**	**0.00%**

Interestingly, when the threshold based denoising method is applied, the findings reveal that the mean ratio for the view-independent and view-dependent voxels at the IFG is 28.2% and 71.8%, respectively (Table 8). Moreover, a small number of view-independent voxels (4.36%) and a majority of view-dependent voxels (95.64%) are observed in the SPL/IPS, although this ratio is smaller than the one is observed in the IFG. On the other hand, it is resulted that the IPL is composed of the view-independent voxels only.

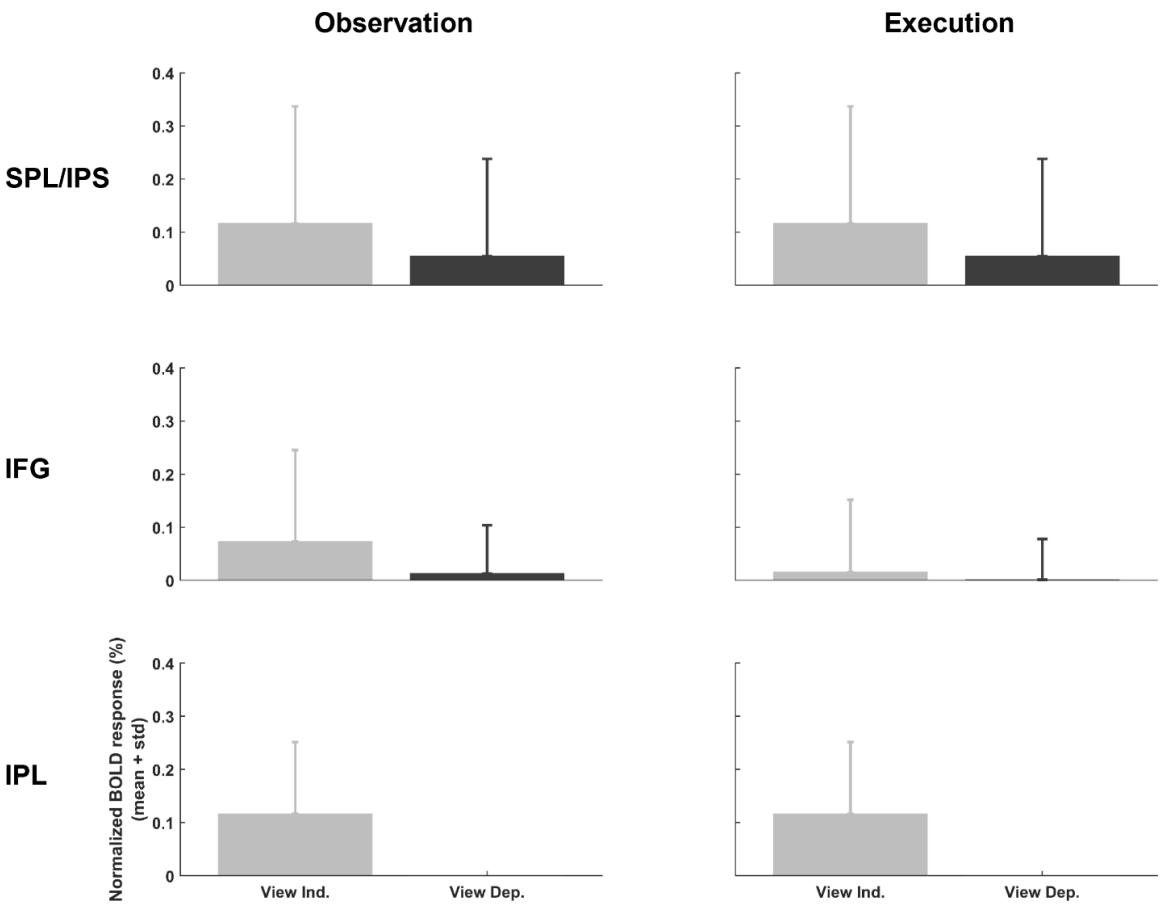


Figure 22. The view-independent and view-dependent BOLD fMRI responses during action observation and execution. The BOLD responses are normalized between 0 and 1.

Finally, the average normalized BOLD fMRI responses of the view-independent and view-dependent voxels are measured (Figure 22). The result reveals that the activities of the view-independent voxels tend to be stronger (from at least 2.12 times in the SPL/IPS up to 13.73 times in the IFG) with larger variance (from at least 1.20 times in the SPL/IPS up to 1.89 times in the IFG) compared to those of the view-dependent voxels, which have weaker responses with smaller variance.

Chapter 5: Summary

Overall, this study proposes a novel neural model that includes a fronto-parietal network to simulate the neural mechanisms underlying the imitation through observational learning by examining its neural dynamics with a parietal visuospatial transformation system during action imitation. Specifically, the fronto-parietal circuit is composed of the IFG (i.e., frontal MNS) and the IPL (i.e., parietal MNS), which are responsible for the visual-to-motor transformation and the sensorimotor predictions, respectively. Moreover, the SPL/IPS is hypothesized to play a critical role in the visuospatial transformations such as re-orientation, rotation, and scaling of the visuospatial representation of the observed actions. As a result, the SPL/IPS can provide the view-independent visuospatial representation to the MNS, which may contribute to coding the intentions inherent in the observed actions (Carr et al., 2003; Fogassi et al., 2005). At the same time, the MT is assumed to provide the view-dependent visual motion representation such as direction and velocity of the action to the MNS (Adelson & Movshon, 1982; Tootell et al., 1995), which may tune the neural activity of the MNS with the view-specific motion information in the observed actions.

The training of these three neural networks (i.e., SPL/IPS, IFG, and IPL) relies on a novel imitation learning strategy that is used to mimic more realistic human behavior in an ecologically valid learning context. First, the neural networks are trained through a continuous cycle of learning by action observation and learning by action execution, which generally constitutes imitation learning. Second, the IFG and the IPL as well as the

SPL/IPS are simultaneously trained using the online update method, where, for the sake of simplicity, the four transformation primitives of the SPL/IPS (i.e., clockwise and counterclockwise rotations as well as translation and scaling) are separately trained in a previous session. Third, the IPL is pre-trained only for a step to simulate the body-babbling, which allows infants to practice their movements through self-generated activity (Meltzoff & Moore, 1997). Interestingly, this learning procedure reveals that the IPL (i.e., sensorimotor predictions) precedes the IFG (i.e., sensorimotor control), which is in agreement with previous findings suggesting the neural processes underlying prediction (i.e., forward computation) and motor control (i.e., inverse computation) during motor learning (Flanagan, Vetter, Johansson, & Wolpert, 2003). This is also consistent with the idea that even an imperfect forward model of the IPL could still be used to train the inverse model of the IFG through fronto-parietal interactions (Jordan & Rumelhart, 1992). As a whole, this newly proposed learning scheme contributes to enhancing the ecological validity of the MNS model during imitation through observational learning.

After learning, this model is assessed under various conditions, where the relative spatial relationships between a demonstrator and an imitator are manipulated. In particular, the conditions are determined by a combination of two explicit variables of y-directional distance (i.e., near, middle, and far) and x-directional distance (i.e., left, center, and right) between these two agents, and two implicit variables of view-angle and anthropometry of them. The results reveal that the proposed model is capable of learning the observed action and reproducing it independently of the differences in anthropometry, distance, viewpoint, and frame of reference between the demonstrator and imitator. Taken as a

whole, through the interaction between the SPL/IPS, IFG, and IPL, this model can reproduce some behavioral and neurophysiological findings from previous mental rotation as well as action imitation studies. Specifically, the model can correctly imitate the observed action regardless of conditions, and the response time for mental transformation monotonically increases with respect to the view angle. Moreover, the BOLD fMRI responses are very similar between action observation and execution. Furthermore, particularly in the simulated BOLD fMRI signal of the IFG, the ratio of view-independent and view-dependent voxels is close to 1:3 as shown in the monkey mirror study (Caggiano et al., 2011), although the nature of the neural signals are different; one is a simulated voxel, and the other is a single neuron recording.

Although the current neural model can imitate through observational learning of actions under various conditions between the demonstrator and the imitator, it has several limitations that are planned to be addressed in future works. In particular, the following three important limitations will be addressed in the near future; i) an incomplete temporo-parieto-frontal circuit due to the simplified STS implementation, ii) the lack of codes of goals in the action, and iii) a more explicit hand modeling for grasping. To overcome these limitations, additional brain regions such as the rPFC, STS, and canonical neuron system in the IFG will be modeled using neural networks, which allow for matching and processing of goals inherent in the grasping action. In conclusion, this novel neural model offers a first step in developing a future computational platform that will allows to further examine the neural mechanisms underlying action observation and imitation by incorporating the MNS, SPL/IPS, and MT. Also, from a practical standpoint, this neural model can contribute to the development of adaptive neuromimetic controllers for

autonomous humanoid robots with potential applications for human-robot interactions, where the robot learns from humans (or from other robots) allowing thus robust and flexible motor performance in ecologically valid situations.

Chapter 6: Discussion

6.1 Mirror system and its view-independent as well as view-dependent activities

This neural model can perform imitation through observational learning independently of the differences in viewpoint between the demonstrator and the imitator, thus resulting in similar imitated kinematics as well as comparable neural activities in the frontal MNS. This is consistent with the studies in nonhuman primates, which have reported that the responses in F5 mirror neurons are similar regardless of the demonstrator's position (Craighero, Metta, Sandini, & Fadiga, 2007; Gallese et al., 1996). This is also in agreement with more recent findings of view-independent and view-dependent F5 mirror neurons, in which the minority (26%) of the F5 mirror neurons are view-independent (Caggiano et al., 2011).

In addition to such monkey studies, some human MNS studies have demonstrated that the frontal MNS (or IFG) is more sensitive to the viewpoint, whereas the parietal MNS (or IPL) is responsive independently of the viewpoint (Aziz-Zadeh et al., 2002; Hétu et al., 2011; Oosterhof et al., 2012). Specifically, the activity of the frontal MNS strongly correlates with the first-person perspective action (i.e., egocentrically transformed representation) rather than the third-person perspective action, whereas the activity of parietal MNS is similar regardless of the differences in viewpoint.

In the current study, the view-independent visual representation of the action is assumed to be provided by the SPL/IPS considering various empirical studies (Andersen, 1987; Buneo & Andersen, 2006; Grefkes & Fink, 2005; Hesse et al., 2009; Oosterhof et al.,

2012; G. Rizzolatti et al., 1998). Particularly, it is hypothesized that the SPL/IPS performs an allocentric-to-egocentric transformation for the observed action, and sends this egocentrically represented action to the frontal MNS to plan the neural commands (i.e., visual-to-motor transformation). Interestingly, this mechanism can effectively address the findings that) the activity of the frontal MNS strongly correlates with the egocentrically transformed representation of action (Oosterhof et al., 2012), and ii) the frontal MNS responds similarly regardless of various spatial relationships between the demonstrator and the imitator (Caggiano et al., 2011; Craighero et al., 2007; Gallese et al., 1996). Under such a mechanism, it is supposed that the view-independent MNS may encode the goals or intentions in the observed actions independently of the details of the visual inputs such as differences in anthropometry, distance, frame of reference, and viewpoint (Caggiano et al., 2011). However, it must be noted that the current model has a limitation in representing the higher level coding of actions since it incorporates the lowest level of coding (i.e., action specific motor states such as kinematics) as well as the low level of coding (i.e., independent of specific action trajectories or states) of actions. In general, it is hypothesized that the STS is required to encode higher level of representation of action; for instance, the STS allows the action to be represented independent of specific end-effectors so that it can code for an action with either hand (Iacoboni, 2005). Moreover, the rPFC needs to be involved to code the highest level of coding for goals in actions, which is also somewhat abstract level that is independent of end-effectors or actions; for instance a reach action to an apple is performed to satisfy hunger (Decety et al., 1997). Although the proposed neural model does not cover these higher and highest levels of coding of intentions in actions, it is still important to examine

the lowest and lower levels of coding in actions to model various experimental evidences in action imitation (Oztop, Kawato, & Arbib, 2013). Furthermore, another recent modeling effort based on this architecture suggests that this approach can be successfully employed to process higher levels of goals (R. J. Gentili et al., 2015).

Besides the view-independent F5 mirror neurons, the view-dependent F5 mirror neurons (74%) are also found in monkeys, where the ratio of view-independent and view-dependent neurons is about 1:3 (Caggiano et al., 2011). Interestingly, in the simulation results of the SPL/IPS and IFG, although the view-dependent voxels are in the majority, the assessment results reveal that their activities are weaker and less variable than the activities of the view-independent voxels. This may indicate that the small portion of the view-independent voxels can have a stronger effect on the imitation in the MNS while they respond to a variety of visual inputs, which lead to higher variance. On the other hand, since the view-dependent voxels are a specialized voxel to a specific visual input, their variance may be small. However they may contribute to the imitation in the MNS due to a large number of neural populations. Beyond the proportion of view-independent neurons to view-dependent neurons, it has been suggested that the view-dependent mirror neurons play two critical roles i) in the formation of view-independent action goals by modulating their associated visual aspects (Caggiano et al., 2011; Logothetis & Sheinberg, 1996; Tanaka, 1996), and ii) in tuning the viewpoint transformation processing by providing feedback signals to the SPL/IPS (Andersen et al., 1997; Wolpert, Goodbody, et al., 1998). In the current model, such a view-dependent MNS is implemented by employing the visual motion representation of the observed action through the MT. Interestingly, the view-dependent representations to the MNS allows the

model to mimic the empirical findings that i) the ratio of the view-independent and view-dependent voxels in the IFG is 1:3 as shown in the monkey study (Caggiano et al., 2011), and ii) the IPL is still similarly responsive to the observed action independently of the condition between the demonstrator and the imitator (Oosterhof et al., 2012). However, it is critical to note that the current neural architecture is not designed to model the neuronal activity of the MNS but rather developed to model the functional roles of the MNS at a neural population level. Therefore, it is important to keep in mind that the results should be carefully investigated and interpreted because the simulated data are assumed to model human mirror system, whereas most of the empirical data are obtained from the mirror neurons of the nonhuman primates.

6.2 Mental transformation and mirror system

It has been suggested that the PPC interacts with the inverse and forward models during action imitation (Andersen et al., 1997; Wolpert, Goodbody, et al., 1998). Specifically, the PPC converts the visuospatial locations of goals inherent in actions into the corresponding motor coordinates, which are subsequently sent to the inverse model (Andersen et al., 1997). Then, the PPC simultaneously updates estimates of the current spatial state of the body by combining sensory feedbacks with the predictive motor commands that are the consequences of the forward process (Wolpert, Goodbody, et al., 1998). Therefore, this suggests that the PPC (particularly SPL/IPS) can provide a useful

framework for gaining new insights into the functional roles of the SPL/IPS in accordance with the MNS requirements during imitation learning.

As such, the visuospatial transformation system of the SPL/IPS could play a critical role in action observation and execution, since it allows the MNS to process both other's actions and self-actions in a single mechanism. In particular, the SPL/IPS is composed of at least four subnetworks performing clockwise and counterclockwise rotations as well as translation and scaling transformation (Burton, Wagner, Lim, & Levy, 1992). Interestingly, the neural processing of the SPL/IPS can reproduce various important neurobehavioral results. Specifically, consistent with other experimental studies, this network can generate a similar linear pattern in neural processing time correlated with response time with respect to the magnitude of the rotation angle but independently of the rotation direction (Bock & Dalecki, 2015; Dalecki et al., 2012). It must be noted that those experimental results are typically obtained during mental rotations of human body parts, complex scene, or objects, which are not directly related with action observation and imitation. Recently, Schwabe et al. (2009) showed an interesting result that the activation timing of the MNS strongly correlates with the reaction time required to perform the mental rotation of human body under various viewpoints (Schwabe, Lenggenhager, & Blanke, 2009). This result implicates that complex mental rotation requires more processing time upstream prior to reaching the MNS, which is consistent with the current results in that the processing time of the SPL/IPS increases linearly with respect to the view angle. As such, by examining the functional relationship of the SPL/IPS and the MNS in action observation and imitation, the proposed neural model can

contribute to the reinforcement of the perception-action coupling in the general framework of the neural simulation of action theory (Jeannerod, 2001).

6.3 Comparison with other MNS models

Besides a novel learning strategy and neurophysiological assessment through a synthetic BOLD fMRI generation model, the main difference between this model and other MNS models is the modeling of a visuospatial transformation system. Particularly, this transformation system is hypothesized to be embedded in the SPL/IPS, and subserves the MNS by providing the view-independent representations of the observed action to the fronto-parietal network. As such, the present work complements previously proposed MNS models, which have not taken into account a neural component that learns various visuospatial transformations between the demonstrator and the imitator (J. Bonaiuto & Arbib, 2010; J. B. Bonaiuto et al., 2007; J. Demiris & Hayes, 2002; Y. Demiris & Johnson, 2003; Oztop & Arbib, 2002; Oztop et al., 2005). In addition, the current neural model shares some common features such as the use of internal model framework with other MNS modeling efforts, even though the network structures and their functional roles are different. Particularly, Demiris and colleagues (J. Demiris & Hayes, 2002; Y. Demiris & Johnson, 2003) as well as Tani and colleagues (Tani, Ito, & Sugita, 2004) proposed a similar approach to the one used in the current study by combining inverse and forward models. However, although an inverse-forward coupling is employed, their studies focus on learning new behaviors to classify observed actions, whereas this study examines the functional role of visuospatial transformation system in imitation through

observational learning. Another difference is that the forward model was not trained through peripheral (e.g., visual) feedback in Demiris's models (2002; 2003). In addition, the self-organization map is employed to encode the distributed representation of actions in Tani's model (2004), whereas the RBF network is used in the current study for the regional representation of actions. Moreover, it must be noted that the current neural model is trained with the supervised learning method, which has a limited ecological validity compared to the unsupervised learning.

Other interesting MNS models have been developed to emphasize the grasp-related mirror activity (Oztop & Arbib, 2002) and to infer the mental states of the demonstrators during action observation (Oztop et al., 2005). Particularly, Oztop & Arbib (2002) focused on the functional roles of the F5 mirror neurons and the F5 canonical neurons in a reaching and grasping action, whereas the current study does not include the canonical neuron system. On the other hand, Oztop and colleagues (2005) strongly emphasized the functional role of the PFC in inferring others' mental states (or intentions) when they perform an action, whereas the current study simply uses the rPFC for a trigger system. It must be noted that the proposed neural model in this study can be complementary to the MSI model, because the intention or goal in actions is assumed to be represented in a view-invariant way, which is provided by the SPL/IPS to the MNS (Carr et al., 2003; Fogassi et al., 2005). At this moment, only few computational MNS studies have accounted, to a limited extent, such a view-invariant representation of the observed action. For instance, Lopes and colleagues (2005) proposed a computational model combining the visuomotor map (i.e., F5 mirror neurons) and the viewpoint transformation for learning by imitation. However, their viewpoint transformation model has no

biological relevance and is rather implemented by a pure transformation matrix. More recently, Arie and colleagues (Arie, Arakaki, Sugano, & Tani, 2012) showed that their MNS model could successfully imitate a novel action albeit the discrepancies in trajectory (up to ±15°) between the observed and executed actions. However, by incorporating the SPL/IPS, the MNS model in this study can process any angular variation. Therefore, the proposed neural architecture allows the MNS components to successfully learn to imitate the observed action with less spatial regulations.

6.4 Synthetic functional neuroimaging data

The generation of the simulated BOLD fMRI data provides a very useful tool to validate the neural activities of the computational model beyond the traditional behavioral assessment by contrasting the simulated and actual brain activities. Specifically, the simulated BOLD fMRI responses from the proposed model are consistent with the findings from other relevant literatures, which have revealed that the IFG responses are similar between action observation and execution. These synthetic BOLD fMRI data reveal the emergence of neural populations that are consistent with the view-independent and view-dependent mirror neurons, which are important features of the mirror neurons. It must be noted that only few modeling studies have employed synthetic brain imaging techniques to help elucidate the relation between simulated neural activity and actual functional brain imaging data (particularly the fMRI data). However, because the hemodynamic measurements reflect synaptic activity rather than neuronal activity, it is likely that inhibitory synaptic activity leads to increased hemodynamic measurements,

even though it actually results in decreased neuronal spiking. Therefore, electrophysiological approaches such as the electrocorticography can be useful to elucidate the knowledge gaps between neuronal activity and the associated changes in CMRO2, CBF, and CBV (Barry Horwitz, 2004). Moreover, due to the limitations of low temporal resolution in fMRI, the simulated functional neuroimaging data in this study cannot reflect the neural dynamics associated with imitation learning. Therefore, a simulation through EEG or MEG technologies, which have higher temporal resolution, could provide a framework for temporal information on imitation tasks. In addition, EEG or MEG data provided by the model could be validated through experiments employing the same type of techniques in a context of ecologically valid cognitive motor performance tasks.

Finally, to validate the predictions made by the current neural model, the following experimental design for the actual fMRI studies is proposed below. Although both block and event-related designs (see Appendix D) can be employed, a mixed design combining these two methods may be appropriate for the validation test (Amaro & Barker, 2006; Chawla, Rees, & Friston, 1999; Petersen & Dubis, 2012). Chawla et al. (1999) initially proposed the mixed designs to examine the effects of selective attention (block factor) that modulates neural activity evoked by two transient stimuli (motion and color stimuli; event-related factor). In accordance with this idea, the proposed validation test can investigate the modulation of evoked responses in frontal and parietal MNS as well as the SPL/IPS by attention to view-angle or distance. Specifically, in some blocks, the stimuli with the view-angle attribute (i.e., left, center, and right) can be are presented, in other blocks, the distance-variant stimuli (i.e., near, middle, and far) can be presented to the

imitator. The stimulus events can last for 1 s and the interstimulus interval (ISI) can be randomly selected from a uniform distribution that ranged from 1 to 15 seconds. Each task block can include 6 trials of stimuli (i.e., 2 repetition of 3 conditions) in random order followed by null stimulus events (i.e., resting condition) for 20 s. These test sequences can be replicated 20 times (i.e., randomly ordered 10 view-angle blocks and 10 distance blocks) to acquire 20 scans per each condition.

6.5 Limitations of the model and future work

Although the current neural model contributes to understanding the underlying mechanisms of the dynamics between the visuospatial transformation and the MNS during action observation and imitation, it also has several limitations that could be addressed in the future. First, from a computational point of view, it is challenging to train three neural networks to imitate the observed actions within the large workspace considered in this study (see section 3.4). Therefore, the entire workspace is currently subdivided into nine smaller local spatial areas (within which the demonstrator is located), where the neural networks are separately trained to bring more efficiency in the computation. However, a single neural network for each of the SPL/IPS, IFG, and IPL could clearly learn the entire workspace if more computational resources are available in the future. Second, the current neural model can perform only two-dimensional planar actions in the horizontal surface. Therefore, future works will focus on achieving a three-dimensional motor control capability to process any action performed in a three-dimensional workspace. Third, the implementations of the rPFC and STS are very

simplified in the current study. A possible solution would be to model these structures by employing neural networks to further explore the coding of the intention that underlies observed actions. Specifically, the PFC could incorporate, but not limited to, inhibitory and working memory mechanisms. Moreover, it must be noted that since the STS has also view-independent (3%) and view-dependent neurons (97%) (Oram & Perrett, 1996), a neural model of this region should contribute to understanding the functional roles of these two types of neurons in the temporo-parieto-frontal network. Such a temporo-parieto-frontal network could be extended by modeling two hemispheres linked by a corpus callosum to examine the relationship between MNS and language as well as MNS and handedness during observation and imitation. Furthermore, other brain regions not currently modeled such as canonical neuron system in the IFG could also be included to allow the imitator to manipulate the target objects (e.g., grasping objects) during action execution (Grèzes, Armony, Rowe, & Passingham, 2003).

Finally, in the future, this work could be extended in two possible ways; i) the first one would be to predict specific dysfunctions observed in neural disorders such as ASD that are assumed to be caused by a failure in the development of the MNS, and ii) the second is to assess this neural model with an actual humanoid robots platform such as Baxter™ (Rethink Robotics, Inc., Boston, MA, USA) to assess specific neural mechanisms underlying action observation and imitation as well as develop applications relevant to human-robot interactions in a real-world environment.

Appendices

Appendix A. Balloon model

The balloon model is a biomechanical model that produces transient dynamics of CBV and HbR to examine their influences on BOLD response (Buxton, Wong, & Frank, 1998; K. J. Friston et al., 2000). Although the model has been refined and extended, its fundamental idea can be expressed as (Buxton et al., 1998):

$$\begin{cases} \frac{dq}{dt} &= \frac{1}{\tau_0}\left(f_{in}(t)\frac{E(t)}{E_0} - f_{out}(v)\frac{q(t)}{v(t)}\right) \\ \frac{dv}{dt} &= \frac{1}{\tau_0}\left(f_{in}(t) - f_{out}(v)\right) \\ E(f) &= 1-(1-E_0)^{\frac{1}{f}} \end{cases} \quad \text{(A1)}$$

where q is the total HbR, v is the volume of the balloon (i.e., CBV), τ_0 is the mean transit time through the venous compartment at rest, f_{in} is the CBF, f_{out} is an increasing function of the balloon volume, E is the net extraction fraction of oxygen, and E_0 is the resting value of E. In Euqation A1, the quantity $f_{in}E/E_0$ indicates the CMRO2 normalized to its value at rest. Therefore, it implies that the balloon model can account for nonlinear neurovascular coupling and the BOLD signal depends on a nonlinear combination of changes in CBF, CBV, and CMRO2 (Buxton, Uludağ, Dubowitz, & Liu, 2004). It must be emphasized that the model can deal with the link between CBF and BOLD signal, that is, the correlates that are not measured with BOLD fMRI (K. J. Friston et al., 2000). In summary, the balloon model is a very powerful tool that can simulate both neurovascular coupling and hemodynamic responses, which in turn lead to the changes in BOLD responses (see Figure 2).

Appendix B. Volterra kernels

A continuous time-invariant system can be expanded in Volterra series as

$$\boldsymbol{y(t) = h_0 + \sum_{n=1}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_n(\tau_1, \cdots, \tau_n) \prod_{k=1}^{n} u(t - \tau_k)\, d\tau_k} \quad \text{(B1)}$$

where $\boldsymbol{y(t)}$ and $\boldsymbol{u(t)}$ are respectively the output and input of the system, and $\boldsymbol{h_n(\tau_1, \cdots, \tau_n)}$ is the n-th-order Volterra kernel. It is clear that the output depends not only on the current input at time $\boldsymbol{\tau_n}$, but also the past input at time $\boldsymbol{\tau_1, \cdots, \tau_{n-1}}$. In other words, the Volterra series can represent dynamic systems with memory. In the context of nonlinear hemodynamic convolution model, the expansion (Equation B1) can be approximated with the first- and second-order Volterra kernels over a finite time constant T representing memory capacity:

$$\begin{aligned} \boldsymbol{y(t)} \approx \quad & \boldsymbol{h_0} + \\ & \boldsymbol{\int_0^T h_1(\tau_1) u(t - \tau_1) d\tau_1} + \\ & \boldsymbol{\int_0^T \int_0^T h_2(\tau_1, \tau_2) u(t - \tau_1) u(t - \tau_2) d\tau_1 d\tau_2} \end{aligned} \quad \text{(B2)}$$

where the integrals start at zero, which reflects that neuronal changes precede hemodynamic responses (Karl J. Friston et al., 1998).

In practical signal processing, the Volterra kernels can be estimated using a relatively small number of temporal basis functions in the discretized time form as:

$$\begin{aligned} \boldsymbol{h_0} &= {}^{0}\boldsymbol{\beta} \\ \boldsymbol{h_1(\tau_1)} &= \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \boldsymbol{x}_i(\boldsymbol{\tau_1}) \\ \boldsymbol{h_2(\tau_1, \tau_2)} &= \sum_{i=1}^{P}\sum_{j=1}^{P} {}^{2}\boldsymbol{\beta}_{ij} \boldsymbol{x}_i(\boldsymbol{\tau_1}) \boldsymbol{x}_j(\boldsymbol{\tau_2}) \end{aligned} \quad \text{(B3)}$$

where all $\boldsymbol{\beta}$ are scaling factors, all $\boldsymbol{x_i(\tau)}$ are temporal basis functions, and $\boldsymbol{P}$ is a small finite number of temporal basis functions. Then, the approximated Volterra expansion (Equation B2) can be expressed with these discretized Volterra kernels (Equation B3) and continuous time convolution formula (See Equation 4). For example, the first-order term of Equation B2 is

$$\begin{aligned} \int_0^T \boldsymbol{h_1(\tau_1) u(t-\tau_1) d\tau_1} &= \int_0^T \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \boldsymbol{x}_i(\boldsymbol{\tau_1})\, \boldsymbol{u(t-\tau_1) d\tau_1} \\ &= \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \int_0^T \boldsymbol{x}_i(\boldsymbol{\tau_1}) \boldsymbol{u(t-\tau_1) d\tau_1} \\ &= \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \cdot \left(\boldsymbol{u(t)} * \boldsymbol{x}_i(\boldsymbol{\tau_1})\right) \\ &= \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \boldsymbol{X}_i(\boldsymbol{t}) \end{aligned} \quad \text{(B4)}$$

where $\boldsymbol{X_i(t)}$ is a set of i-th response function. Therefore, using a similar approach in Equation B4, the Volterra expansion can be expressed within GLM (Karl J. Friston et al., 1998):

$$\boldsymbol{y(t)} = {}^{0}\boldsymbol{\beta} + \sum_{i=1}^{P} {}^{1}\boldsymbol{\beta}_i \boldsymbol{X}_i(\boldsymbol{t}) + \sum_{i=1}^{P}\sum_{j=1}^{P} {}^{2}\boldsymbol{\beta}_{ij} \boldsymbol{X}_i(\boldsymbol{t}) \boldsymbol{X}_j(\boldsymbol{t}) + \boldsymbol{\varepsilon(t)} \quad \text{(B5)}$$

where $\boldsymbol{\varepsilon(t)}$ is an error term. Then, three explanatory variables, 1, $\boldsymbol{X_i(t)}$, and $\boldsymbol{X_i(t)X_j(t)}$, constitute the columns of the design matrix.

Appendix C. Matrix Similarity

Two $\boldsymbol{n} \times \boldsymbol{n}$ square matrices X and Y are said to be similar if there exists a nonsingular (or invertible) $\boldsymbol{n} \times \boldsymbol{n}$ square matrix P such that

$$\mathrm{Y} = \mathrm{P}^{-\mathbf{1}}\mathrm{XP} \tag{C1}$$

where the transformation $\mathrm{X} \mapsto \mathrm{P}^{-\mathbf{1}}\mathrm{XP}$ is called as a similarity transformation or conjugation of the matrix X. Particularly, similar matrices have a critical property in the theorem that if two square matrices X and Y are similar, then they have the same characteristic polynomial and thus the same eigenvalues. This can be proved using Equation C1 such that

$$\begin{aligned} \boldsymbol{det}(\mathrm{Y} - \boldsymbol{\lambda}\mathrm{I}) &= \boldsymbol{det}\left(\mathrm{P}^{-\mathbf{1}}\mathrm{XP} - \boldsymbol{\lambda}\mathrm{I}\right) \\ &= \boldsymbol{det}\left(\mathrm{P}^{-\mathbf{1}}(\mathrm{X} - \boldsymbol{\lambda}\mathrm{I})\mathrm{P}\right) \\ &= \boldsymbol{det}\left(\mathrm{P}^{-\mathbf{1}}\right)\boldsymbol{det}(\mathrm{X} - \boldsymbol{\lambda}\mathrm{I})\boldsymbol{det}(\mathrm{P}) \\ &= \boldsymbol{det}(\mathrm{X} - \boldsymbol{\lambda}\mathrm{I}) \end{aligned} \tag{C2}$$

Moreover, the determinant function is used to find a matrix's eigenvalues. For a general $\boldsymbol{n} \times \boldsymbol{n}$ matrix A, the characteristic polynomial of the matrix A is defined by

$$\boldsymbol{p}_{\mathrm{A}}(\boldsymbol{\lambda}) = \boldsymbol{det}(\boldsymbol{\lambda}\mathrm{I} - \mathrm{A}) \tag{C3}$$

where the roots of this equation are the eigenvalues of its associated matrix A if and only if there is an eigenvector $\mathrm{v} \neq 0$ such that $\mathrm{Av} = \boldsymbol{\lambda}\mathrm{v}$ or $(\boldsymbol{\lambda}\mathrm{I} - \mathrm{A})\mathrm{v} = 0$. Therefore, there exists up to $\boldsymbol{n}$ distinct orthonormal eigenvalues $\boldsymbol{\lambda_i}$ ($1 \leq \boldsymbol{i} \leq \boldsymbol{n}$).

Appendix D. Experimental designs in neuroimaging studies

The time between the onset of successive trials (or stimuli), which is often referred to as the stimulus onset asynchrony (SOA) or intertrial interval (ITI), correlates with the scaling and the shape of the BOLD responses (Figure D1).

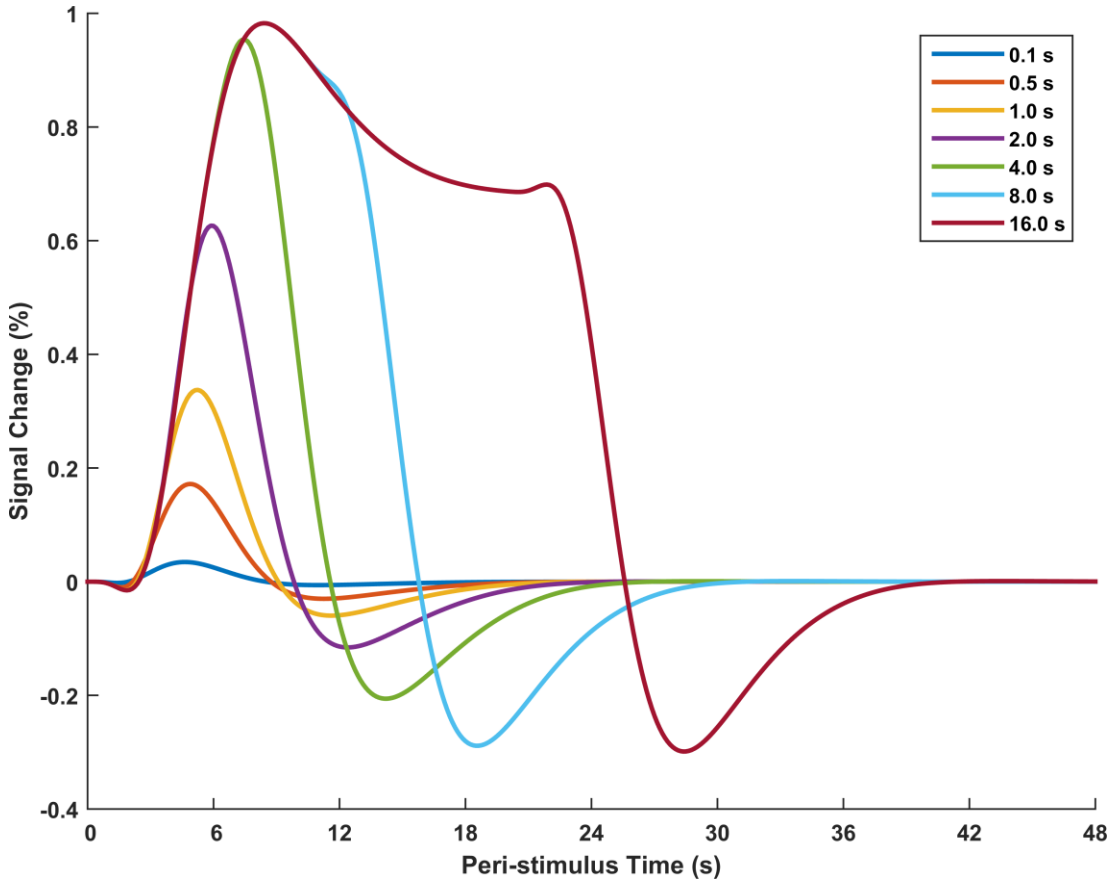


Figure D1. HR with respect to stimulus duration. All signals are presented at time 0 with discrete durations (from 0.1 to 16.0 s, depicted with different colors). For the relatively short stimuli (< 4.0 s), the differences are observed in the scaling of the predicted BOLD responses with little influence on their shapes. The prolonged stimuli (> 8.0 s) cause the major differences in the shape of the BOLD responses rather than in the scaling of them. These differences lead to the different BOLD response patterns observed in the block designs and in the event-related designs.

Specifically, relatively shorter trials (i.e., 4-s or less, but usually with zero duration) have influence on the scaling of the BOLD responses rather than their shape (R. Henson, 2007). This phenomenon is effectively modeled in the event-related designs, in which discrete multiple trials with short duration are randomly presented one at a time so that they can cause brief bursts of neural activity. On the other hand, for relatively longer trials, the BOLD responses have a plateau, which is related to the shape of the BOLD responses (R. Henson, 2007). Such a pattern in BOLD responses is better reflected with the block designs, in which a series of same conditional stimuli is consecutively presented over a period of time so that it can bring sustained neural activity. Interestingly, in most brain imaging studies, these two types of experimental designs have been mainly adopted. As noted, they are clearly distinguished by the stimulus presentation scheme, and each has its strong and weak points. Regardless of the experimental designs, the BOLD signals measured during one condition are then compared with other BOLD signals of different task conditions.

D.1 Block design

In block designs, a test consists of a series of blocks lasting for a specific time (e.g., 30 s/block). In each block, participants repeatedly perform the same tasks while the same stimuli are subsequently presented to the subjects in accordance with the stimulus duration (e.g., 4 s) and the interstimulus interval (ISI4; e.g., 10 s). Generally, two types of blocks are used in such a way that the task blocks, where the subject performs a particular task, alternate with the resting blocks, in which the subjects simply rest. A majority of

[4] Generally, for block designs, SOA = ISI + stimulus duration, whereas for event-related designs, SOA = ISI.

studies has measured neuroimaging data using block designs particularly for higher SNR, straightforward statistical analysis, and greater likelihood of detecting a neural response to the stimuli (Amaro & Barker, 2006). However, the subjects are more likely to show anticipatory responses to the stimuli.

D.2 Event-related design

In event-related designs, a test is composed of several discrete trials. Each trial is repeated a specific number of times (e.g., 20 times/trial) in such a way that it is presented in random order over the entire test. This design has been typically employed to capture dynamic neural activity with relatively high temporal resolution during simple cognitive, sensory, or motor tasks. However, there is a potential issue in complex tasks, in which much of the task-unrelated neural activity tends to be indirectly induced under the same experimental control. Therefore, this can often lead to incorrect results related to different internal processing.

Incidentally, two types of event-related designs have been mainly proposed considering ISI, that is, slow and rapid event-related designs (Amaro & Barker, 2006). Specifically, the slow type uses relatively long ISIs (e.g., 15 s with random delays or jittering) between successive trials, while the rapid type includes relatively short ISIs (e.g., 3 s with jittering). Although the short ISIs result in overlapped BOLD responses that need to be separated in analysis, recent studies mostly use the rapid designs due to two weaknesses in the slow designs (Amaro & Barker, 2006; Ashby, 2015). First, some task-unrelated cognitive processes, which emerge during these long ISIs, are likely to contaminate the induced BOLD responses. Second, assuming a fixed-duration test, the long ISIs reduce

the number of trials in the test, thus it decreases the statistical power in analysis. Regardless of the types, the event-related designs are suitable for behaviorally more realistic tasks with various experiment conditions, post-hoc analysis, and examining BOLD responses to individual trials (Amaro & Barker, 2006).

Bibliography

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*(5892), 523–525. http://doi.org/10.1038/300523a0

Amaro, E. J., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition*, *60*(3), 220–232. http://doi.org/10.1016/j.bandc.2005.11.009

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* (1st ed., pp. 483–518). Bethesda, MD: American Physiological Society. http://doi.org/10.1002/cphy.cp010512

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, *20*, 303–330. http://doi.org/10.1146/annurev.neuro.20.1.303

Arbib, M. A., Billard, A., Iacoboni, M., & Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks*, *13*(8-9), 975–997. http://doi.org/10.1016/S0893-6080(00)00070-8

Arbib, M. A., Bischoff, A., Fagg, A. H., & Grafton, S. T. (1994). Synthetic PET: Analyzing Large-Scale Properties of Neural Networks. *Human Brain Mapping*, *2*(4), 225–233. http://doi.org/10.1002/hbm.460020404

Arbib, M. A., Fagg, A. H., & Grafton, S. T. (2002). Synthetic PET Imaging for Grasping : From Primate Neurophysiology to Human Behavior. In F. T. Sommer & A. Wichert (Eds.), *Exploratory Analysis and Data Modeling in Functional Neuroimaging* (1st ed., pp. 231–250). Cambridge, MA: The MIT Press.

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*(5), 729–741. http://doi.org/10.1016/j.robot.2011.11.005

Ashby, F. G. (2015). An Introduction to fMRI. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An Introduction to Model-Based Cognitive Neuroscience* (1st ed., pp. 91–112). New York, NY: Springer-Verlag New York. http://doi.org/10.1007/978-1-4939-2236-9_5

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the Human Mirror Neuron System. *The Journal of Neuroscience*,

26(11), 2964–2970. http://doi.org/10.1523/JNEUROSCI.2921-05.2006

Aziz-Zadeh, L., Maeda, F., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2002). Lateralization in motor facilitation during action observation: a TMS study. *Experimental Brain Research*, *144*(1), 127–131. http://doi.org/10.1007/s00221-002-1037-5

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. http://doi.org/10.1016/j.humov.2015.01.017

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*(4), 341–359. http://doi.org/10.1007/s00422-010-0371-0

Bonaiuto, J. B., Rosta, E., & Arbib, M. A. (2007). Extending the mirror neuron system model, I. Audible actions and invisible grasps. *Biological Cybernetics*, *96*(1), 9–38. http://doi.org/10.1007/s00422-006-0110-8

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1. *The Journal of Neuroscience*, *16*(13), 4207–4221.

Brandauer, B., Timmann, D., Häusler, A., & Hermsdörfer, J. (2010). Influences of Load Characteristics on Impaired Control of Grip Forces in Patients With Cerebellar Damage. *Journal of Neurophysiology*, *103*(2), 698–708. http://doi.org/10.1152/jn.00337.2009

Broomhead, D. S., & Lowe, D. (1988a). Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, *2*(3), 321–355.

Broomhead, D. S., & Lowe, D. (1988b). *Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks*. London, UK.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., … Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*(2), 400–404. http://doi.org/10.1046/j.1460-9568.2001.01385.x

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., & Rizzolatti, G. (2004). Neural Circuits Underlying Imitation Learning of Hand Actions: An Event-Related fMRI Study. *Neuron*, *42*, 323–334. http://doi.org/10.1016/S0896-6273(04)00181-3

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186–198. http://doi.org/10.1038/nrn2575

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. http://doi.org/10.1016/S0893-6080(98)00109-9

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. http://doi.org/10.1037/0033-295X.95.1.49

Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A Self-Organizing Neural Model of Motor Equivalent Reaching and Tool Use by a Multijoint Arm. *Journal of Cognitive Neuroscience, 5*(4), 408–435. http://doi.org/10.1162/jocn.1993.5.4.408

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. http://doi.org/10.1016/j.neuropsychologia.2005.10.011

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *TRENDS in Cognitive Sciences, 11*(7), 290–298. http://doi.org/10.1016/j.tics.2007.05.004

Burton, L. A., Wagner, N., Lim, C., & Levy, J. (1992). Visual Field Differences for Clockwise and Counterclockwise Mental Rotation. *Brain and Cognition, 18*(2), 192–207. http://doi.org/10.1016/0278-2626(92)90078-Z

Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage, 23*(Suppl. 1), S220–S233. http://doi.org/10.1016/j.neuroimage.2004.07.013

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of Blood Flow and Oxygenation Changes During Brain Activation: The Balloon Model. *Magnetic Resonance in Medicine, 39*(6), 855–864. http://doi.org/10.1002/mrm.1910390602

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., & Casile, A. (2011). View-Based Encoding of Actions in Mirror Neurons of Area F5 in Macaque Premotor Cortex. *Current Biology, 21*(2), 144–148. http://doi.org/10.1016/j.cub.2010.12.022

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. http://doi.org/10.1073/pnas.0935845100

Carroll, W. R., & Bandura, A. (1987). Translating Cognition Into Action: The Role of Visual Guidance in Observational Learning. *Journal of Motor Behavior*, *19*(3), 385–398.

Chaminade, T., Meltzoff, A. N., & Decety, J. (2005). An fMRI study of imitation: action representation and body schema. *Neuropsychologia*, *43*(1), 115–27. http://doi.org/10.1016/j.neuropsychologia.2004.04.026

Chao, L. L., & Martin, A. (2000). Representation of Manipulable Man-Made Objects in the Dorsal Stream. *NeuroImage*, *12*(4), 478–84. http://doi.org/10.1006/nimg.2000.0635

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671–676. http://doi.org/10.1038/10230

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. http://doi.org/10.1109/72.80341

Cochin, S., Barthelemy, C., Lejeune, B., Roux, S., & Martineau, J. (1998). Perception of motion and qEEG activity in human adults. *Electroencephalography and Clinical Neurophysiology*, *107*(4), 287–295. http://doi.org/10.1016/S0013-4694(98)00071-6

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., … Belliveau, J. W. (1996). Changes in cortical activity during mental rotation: A mapping study using functional MRI. *Brain*, *119*(1), 89–100. http://doi.org/10.1093/brain/119.1.89

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A Common Network of Functional Areas for Attention and Eye Movements. *Neuron*, *21*(4), 761–773. http://doi.org/10.1016/S0896-6273(00)80593-0

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270. http://doi.org/10.1016/S1053-8119(03)00049-1

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3),

162–173. http://doi.org/10.1006/cbmr.1996.0014

Craighero, L., Metta, G., Sandini, G., & Fadiga, L. (2007). The mirror-neurons system: data and models. *Progress in Brain Research*, *164*, 39–59. http://doi.org/10.1016/S0079-6123(07)64003-5

Crosson, B., Ford, A., McGregor, K. M., Meinzer, M., Cheshkov, S., Li, X., … Briggs, R. W. (2010). Functional Imaging and Related Techniques: An Introduction for Rehabilitation Researchers. *Journal of Rehabilitation Research and Development*, *47*(2), vii–xxxiv. http://doi.org/10.1682/JRRD.2010.02.0017

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology*, *11*(2), 157–163. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11301234

Dalecki, M., Hoffmann, U., & Bock, O. (2012). Mental rotation of letters, body parts and complex scenes: Separate or common mechanisms? *Human Movement Science*. http://doi.org/10.1016/j.humov.2011.12.001

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), 28–30. http://doi.org/10.1038/nn1611

David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage*, *20*(3), 1743–1755. http://doi.org/10.1016/j.neuroimage.2003.07.015

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., … Fazio, F. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, *120*(10), 1763–1777. http://doi.org/10.1093/brain/120.10.1763

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 (Eds.), *Imitation in Animals and Artifacts* (1st ed., Vol. 21, pp. 327–361). Cambridge, MA: The MIT Press. Retrieved from http://www.journals.cambridge.org/abstract_S0263574702224649

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. http://doi.org/10.1080/09540090310001655129

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton,

S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. *Nature Neuroscience*, *2*(6), 563–7. http://doi.org/10.1038/9219

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. http://doi.org/10.1007/BF00230027

Diedrichsen, J., Criscimagna-Hemminger, S. E., & Shadmehr, R. (2007). Dissociating Timing and Coordination as Functions of the Cerebellum. *Journal of Neuroscience*, *27*(23), 6291–6301. http://doi.org/10.1523/JNEUROSCI.0061-07.2007

Dominey, P. F., & Arbib, M. A. (1992). A Cortico-Subcortical Model for Generation of Spatially Accurate Sequential Saccades. *Cerebral Cortex*, *2*(2), 153–175. http://doi.org/10.1093/cercor/2.2.153

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*(1), 103–109. http://doi.org/10.1016/S0926-6410(99)00029-4

Ebadzadeh, M., Tondu, B., & Darlot, C. (2005). Computation of inverse functions in a model of cerebellar and reflex pathways allows to control a mobile mechanical segment. *Neuroscience*, *133*(1), 29–49. http://doi.org/10.1016/j.neuroscience.2004.09.048

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor Facilitation During Action Observation: A Magnetic Stimulation Study. *Journal of Neurophysiology*, *73*(6), 2608–2611. http://doi.org/10.1.1.299.4524

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*(8), 1703–1714. http://doi.org/10.1046/j.1460-9568.2003.02601.x

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., & Suomi, S. J. (2006). Neonatal Imitation in Rhesus Macaques. *PLoS Biology*, *4*(9), 1501–1508. http://doi.org/10.1371/journal.pbio.0040302

Flanagan, J. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction Precedes Control in Motor Learning. *Current Biology*, *13*(2), 146–150. http://doi.org/10.1016/S0960-9822(03)00007-1

Flynn, E., & Whiten, A. (2008). Cultural Transmission of Tool Use in Young Children: A Diffusion Chain Study. *Social Development*, *17*(3), 699–718.

http://doi.org/10.1111/j.1467-9507.2007.00453.x

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. http://doi.org/10.1126/science.1106138

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 (Eds.), *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (1st ed., Vol. 1404, pp. 293–312). Berlin, Germany: Springer-Verlag Berlin Heidelberg. http://doi.org/10.1007/3-540-69342-4_14

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Human Brain Mapping*, *2*(4), 189–210. http://doi.org/10.1002/hbm.460020402

Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear Event-Related Responses in fMRI. *Magnetic Resonance in Medicine*, *39*(1), 41–52. http://doi.org/10.1002/mrm.1910390109

Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics. *NeuroImage*, *12*(4), 466–477. http://doi.org/10.1006/nimg.2000.0630

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593–609. http://doi.org/10.1093/brain/119.2.593

Gangitano, M., Mottaghy, F. M., & Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, *12*(7), 1489–92. http://doi.org/10.1097/00001756-200105250-00038

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD Activity during Mental Rotation and Viewpoint-Dependent Object Recognition. *Neuron*, *34*(1), 161–171. http://doi.org/10.1016/S0896-6273(02)00622-0

Gentili, R., Han, C. E., Schweighofer, N., & Papaxanthis, C. (2010). Motor Learning Without Doing: Trial-by-Trial Improvement in Motor Performance During Mental Training. *Journal of Neurophysiology*, *104*(2), 774–783. http://doi.org/10.1152/jn.00257.2010

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. http://doi.org/10.1007/s12369-014-0276-5

Gentili, R. J., Oh, H., Molina, J., & Contreras-Vidal, J. L. (2011). Neural Network Models for Reaching and Dexterous Manipulation in Humans and Anthropomorphic Robotic Systems. In V. Cutsuridis, A. Hussain, & J. G. Taylor (Eds.), *Perception-Action Cycle* (1st ed., pp. 187–217). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4419-1452-1_6

Gentili, R. J., Papaxanthis, C., Ebadzadeh, M., Eskiizmirliler, S., Ouanezar, S., & Darlot, C. (2009). Integration of Gravitational Torques in Cerebellar Pathways Allows for the Dynamic Inverse Computation of Vertical Pointing Movements of a Robot Arm. *PLoS ONE*, *4*(4), e5176. http://doi.org/10.1371/journal.pone.0005176

Glover, G. H. (1999). Deconvolution of Impulse Response in Event-Related BOLD fMRI. *NeuroImage*, *9*(4), 416–429. http://doi.org/10.1006/nimg.1998.0419

Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D. L., … Andreasen, N. C. (1998). Functional MRI Statistical Software Packages: A Comparative Analysis. *Human Brain Mapping*, *6*(2), 73–84. http://doi.org/10.1002/(SICI)1097-0193(1998)6:2<73::AID-HBM1>3.0.CO;2-H

Goldenberg, G. (2003). Apraxia and Beyond: Life and Work of Hugo Liepmann. *Cortex, 39*(3), 509–524. http://doi.org/10.1016/S0010-9452(08)70261-2

Goldenberg, G. (2009). Apraxia and the parietal lobes. *Neuropsychologia*, *47*(6), 1449–1459. http://doi.org/10.1016/j.neuropsychologia.2008.07.014

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, *21*(2), 215. http://doi.org/10.2307/1268518

Gomi, H., & Kawato, M. (1992). Adaptive feedback control models of the vestibulocerebellum and spinocerebellum. *Biological Cybernetics*, *68*(2), 105–114. http://doi.org/10.1007/BF00201432

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. http://doi.org/10.1007/BF00227183

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3–17.

http://doi.org/10.1111/j.1469-7580.2005.00426.x

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 253–258. http://doi.org/10.1073/pnas.0135058100

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. http://doi.org/10.1016/S1053-8119(03)00042-9

Guenther, F. H., & Barreca, D. M. (1997). Neural Models for Flexible Control of Redundant Systems. In P. G. Morasso & V. Sanguineti (Eds.), *Self-Organization, Computational Maps, and Motor Control* (1st ed., Vol. Advances i, pp. 383–421). Amsterdam, The Netherlands: Elsevier Science B.V.

Guenther, F. H., & Ghosh, S. S. (2003). A Model of Cortical and Cerebellar Function in Speech. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 169–174). Barcelona, Spain: International Phonetic Association. Retrieved from papers://e7d065ae-9998-4287-8af0-c9fa85af8e96/Paper/p41035

Hadjikhani, N., Joseph, R. M., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical Differences in the Mirror Neuron System and Social Cognition Network in Autism. *Cerebral Cortex*, *16*(9), 1276–1282. http://doi.org/10.1093/cercor/bhj069

Hamilton, A. F. de C. (2008). Emulation and mimicry for social interaction: A theoretical approach to imitation in autism. *The Quarterly Journal of Experimental Psychology*, *61*(1), 101–115. http://doi.org/10.1080/17470210701508798

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(25), 15061–15065. http://doi.org/10.1073/pnas.95.25.15061

Haruno, M., Wolpert, D. M., & Kawato, M. (2001). MOSAIC model for Sensorimotor Learning and Control. *Neural Computation*, *13*(10), 2201–20. http://doi.org/10.1162/089976601750541778

Henson, R. (2007). Efficient Experimental Design for fMRI. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (1st ed., pp. 193–210). London, UK: Academic Press. http://doi.org/10.1016/B978-012372560-8/50015-2

Henson, R., & Friston, K. (2007). Convolution Models for fMRI. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (1st ed., pp. 178–192). London, UK: Academic Press. http://doi.org/10.1016/B978-012372560-8/50014-0

Henson, R., Rugg, M. D., & Friston, K. J. (2001). The choice of basis functions in event-related fMRI. *NeuroImage*, *13*(6), S149. http://doi.org/10.1016/S1053-8119(01)91492-2

Herdin, M., Czink, N., Ozcelik, H., & Bonek, E. (2005). Correlation Matrix Distance, a Meaningful Measure for Evaluation of Non-Stationary MIMO Channels. In *2005 IEEE 61st Vehicular Technology Conference* (Vol. 1, pp. 136–140). Piscataway, NJ: IEEE. http://doi.org/10.1109/VETECS.2005.1543265

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. http://doi.org/10.1162/jocn.2009.21058

Hétu, S., Mercier, C., Eugène, F., Michon, P.-E., & Jackson, P. L. (2011). Modulation of Brain Activity during Action Observation: Influence of Perspective, Transitivity and Meaningfulness. *PLoS ONE*, *6*(9), e24728. http://doi.org/10.1371/journal.pone.0024728

Horwitz, B. (2004). Relating fMRI and PET Signals to Neural Activity by Means of Large-Scale Neural Models. *Neuroinformatics*, *2*(2), 251–266. http://doi.org/10.1385/NI:2:2:251

Horwitz, B., Friston, K. J., & Taylor, J. G. (2000). Neural modeling and functional brain imaging: an overview. *Neural Networks*, *13*(8-9), 829–46. http://doi.org/10.1016/s0893-6080(00)00062-9

Horwitz, B., & Tagamets, M.-A. (1999). Predicting Human Functional Maps With Neural Net Modeling. *Human Brain Mapping*, *8*(2-3), 137–142. http://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<137::AID-HBM11>3.0.CO;2-B

Horwitz, B., Tagamets, M.-A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences*, *3*(3), 91–98. http://doi.org/http://dx.doi.org/10.1016/S1364-6613(99)01282-6

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, *15*(6), 632–637. http://doi.org/10.1016/j.conb.2005.10.010

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.-C., … Rizzolatti, G. (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. http://doi.org/10.1073/pnas.241474598

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical Mechanisms of Human Imitation. *Science*, *286*(5449), 2526–2528. http://doi.org/10.1126/science.286.5449.2526

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., … Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, *403*(6766), 192–5. http://doi.org/10.1038/35003194

Imazu, S., Sugio, T., Tanaka, S., & Inui, T. (2007). Differences Between Actual and Imagined Usage of Chopsticks: an fMRI Study. *Cortex*, *43*(3), 301–307. http://doi.org/10.1016/S0010-9452(08)70456-8

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. http://doi.org/10.1007/BF00199471

Jeannerod, M. (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition. *NeuroImage*, *14*(1), S103–S109. http://doi.org/10.1006/nimg.2001.0832

Jordan, M. I., & Rumelhart, D. E. (1992). Forward Models: Supervised Learning with a Distal Teacher. *Cognitive Science*, *16*(3), 307–354. http://doi.org/10.1207/s15516709cog1603_1

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*(6), 718–727. http://doi.org/10.1016/S0959-4388(99)00028-8

Kawato, M., & Gomi, H. (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biological Cybernetics*, *68*(2), 95–103. http://doi.org/10.1007/BF00201431

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science*, *297*(5582), 846–848. http://doi.org/10.1126/science.1070311

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4 Pt 1), 2425–2438. http://doi.org/10.1121/1.417951

Kujala, J., Pammer, K., Cornelissen, P., Roebroeck, A., Formisano, E., & Salmelin, R. (2007). Phase Coupling in a Cerebro-Cerebellar Network at 8-13 Hz during Reading. *Cerebral Cortex*, *17*(6), 1476–1485. http://doi.org/10.1093/cercor/bhl059

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. http://doi.org/10.1038/nature06976

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual Object Recognition. *Annual Review of Neuroscience*, *19*(1), 577–621. http://doi.org/10.1146/annurev.ne.19.030196.003045

Lopes, M., & Santos-Victor, J. (2005). Visual Learning by Imitation With Motor Representations. In *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* (Vol. 35, pp. 438–449). IEEE. http://doi.org/10.1109/TSMCB.2005.846654

Maeda, F., Kleiner-Fisman, G., & Pascual-Leone, A. (2002). Motor Facilitation While Observing Hand Actions: Specificity of the Effect and Role of Observer's Orientation. *Journal of Neurophysiology*, *87*(3), 1329–1335. http://doi.org/10.1152/jn.00773.2000.

Maschke, M., Gomez, C. M., Ebner, T. J., & Konczak, J. (2003). Hereditary Cerebellar Ataxia Progressively Impairs Force Adaptation During Goal-Directed Arm Movements. *Journal of Neurophysiology*, *91*(1), 230–238. http://doi.org/10.1152/jn.00557.2003

Mazzoni, P., & Krakauer, J. W. (2006). An Implicit Plan Overrides an Explicit Strategy during Visuomotor Adaptation. *Journal of Neuroscience*, *26*(14), 3642–3645. http://doi.org/10.1523/JNEUROSCI.5317-05.2006

McDonough, L., Mandler, J. M., McKee, R. D., & Squire, L. R. (1995). The deferred imitation task as a nonverbal measure of declarative memory. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(16), 7580–7584. http://doi.org/10.1073/pnas.92.16.7580

McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., & Horwitz, B. (1994). Network Analysis of Cortical Visual Pathways Mapped with PET. *The Journal of Neuroscience*, *14*(2), 655–666.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn Infants Imitate Adult Facial Gestures. *Child Development*, *54*(3), 702–709. http://doi.org/10.2307/1130058

Meltzoff, A. N., & Moore, M. K. (1997). Explaining Facial Imitation: A Theoretical Model. *Early Development & Parenting*, *6*, 179–192. http://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Meltzoff, A. N., & Moore, M. M. K. (1977). Imitation of Facial and Manual Gestures by

Human Neonates. *Science*, *198*(4312), 74–78. http://doi.org/10.1126/science.897687

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. http://doi.org/10.1097/01.wnr.0000098751.87269.77

Miall, R. C., Reckess, G. Z., & Imamizu, H. (2001). The cerebellum coordinates eye and hand tracking movements. *Nature Neuroscience*, *4*(6), 638–644. http://doi.org/10.1038/88465\r88465 [pii]

Miller, A. J. (1984). Selection of Subsets of Regression Variables. *Journal of the Royal Statistical Society. Series A (General)*, *147*(3), 389–425. http://doi.org/10.2307/2981576

Monaco, S., Cavina-Pratesi, C., Sedda, A., Fattori, P., Galletti, C., & Culham, J. C. (2011). Functional magnetic resonance adaptation reveals the involvement of the dorsomedial stream in hand orientation for grasping. *Journal of Neurophysiology*, *106*(5), 2248–2263. http://doi.org/10.1152/jn.01069.2010

Mulliken, G. H., Musallam, S., & Andersen, R. A. (2008). Forward estimation of movement state in posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(24), 8170–7. http://doi.org/10.1073/pnas.0802602105

Nasrallah, I., & Dubroff, J. (2013). An Overview of PET Neuroimaging. *Seminars in Nuclear Medicine*, *43*(6), 449–461. http://doi.org/10.1053/j.semnuclmed.2013.06.003

Nowak, D. A., Topka, H., Timmann, D., Boecker, H., & Hermsdörfer, J. (2007). The role of the cerebellum for predictive control of grasping. *The Cerebellum*, *6*(1), 7–17. http://doi.org/10.1080/14734220600776379

Nunez, P. L. (1989). Generation of Human EEG by a Combination of Long and Short Range Neocortical Interactions. *Brain Topography*, *1*(3), 199–215. http://doi.org/10.1007/BF01129583

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, *24*(2), 190–8. http://doi.org/10.1016/j.cogbrainres.2005.01.014

Ogawa, S., Lee, T. M., R., K. A., & W., T. D. (1990). Brain magnetic resonance imaging

with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–72. http://doi.org/10.1073/pnas.87.24.9868

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 the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4183–4186). Boston, MD, USA: IEEE. http://doi.org/10.1109/IEMBS.2011.6091038

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 the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2551–2554). San Diego, CA, USA: IEEE. http://doi.org/10.1109/EMBC.2012.6346484

Ollinger, J. M., & Fessler, J. A. (1997). Positron-Emission Tomography. *IEEE Signal Processing Magazine*, *14*(1), 43–55. http://doi.org/10.1109/79.560323

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. http://doi.org/10.1162/jocn_a_00195

Oram, M. W., & Perrett, D. I. (1996). Integration of Form and Motion in the Anterior Superior Temporal Polysensory Area (STPa) of the Macaque Monkey. *Journal of Neurophysiology*, *76*(1), 109–129. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8836213

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. http://doi.org/10.1109/SBRN.1998.730989

Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, *87*(2), 116–140. http://doi.org/10.1007/s00422-002-0318-1

Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, *19*(3), 254–271. http://doi.org/10.1016/j.neunet.2006.02.002

Oztop, E., Kawato, M., & Arbib, M. A. (2013). Mirror neurons: functions, mechanisms

and models. *Neuroscience Letters*, *540*, 43–55. http://doi.org/10.1016/j.neulet.2012.10.005

Oztop, E., Wolpert, D. M., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, *22*(2), 129–151. http://doi.org/10.1016/j.cogbrainres.2004.08.004

Patuzzo, S., Fiaschi, A., & Manganotti, P. (2003). Modulation of motor cortex excitability in the left hemisphere during action observation: a single- and paired-pulse transcranial magnetic stimulation study of self- and non-self-action observation. *Neuropsychologia*, *41*(9), 1272–1278. http://doi.org/10.1016/S0028-3932(02)00293-2

Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *NeuroImage*, *62*(2), 1177–1184. http://doi.org/10.1016/j.neuroimage.2011.09.084

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. http://doi.org/10.1016/S0166-2236(98)01260-0

Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience*, *27*, 169–192. http://doi.org/10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. http://doi.org/10.1016/0926-6410(95)00038-0

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. http://doi.org/10.1038/35090060

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalography and Clinical Neurophysiology*, *106*(4), 283–296. http://doi.org/10.1016/S0013-4694(98)00022-4

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. http://doi.org/10.1093/brain/121.5.815

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9388–9393. http://doi.org/10.1073/pnas.1001791107

Schmidt, R. A. (1975). A Schema Theory of Discrete Motor Skill Learning.

Psychological Review, *82*(4), 225–260. http://doi.org/10.1037/h0076770

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. http://doi.org/10.1002/hbm.20764

Schweighofer, N., Arbib, M. A., & Kawato, M. (1998). Role of the cerebellum in reaching movements in humans. I. Distributed inverse dynamics control. *The European Journal of Neuroscience*, *10*(1), 86–94. http://doi.org/doi:10.1046/j.1460-9568.1998.00006.x

Seitz, R. J., Canavan, A. G. M., Yágüez, L., Herzog, H., Teilmann, L., Knorr, U., … Hömberg, V. (1994). Successive roles of the cerebellum and premotor cortices in trajectorial learning. *NeuroReport*, *5*(18), 2541–2544. http://doi.org/10.1097/00001756-199412000-00034

Shmuelof, L., & Zohary, E. (2007). Watching Others' Actions: Mirror Representations in the Parietal Cortex. *The Neuroscientist*, *13*(6), 667–672. http://doi.org/10.1177/1073858407302457

Shmuelof, L., & Zohary, E. (2008). Mirror-image representation of action in the anterior parietal cortex. *Nature Neuroscience*, *11*(11), 1267–1269. http://doi.org/10.1038/nn.2196

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. http://doi.org/10.1126/science.273.5281.1564

Southgate, V., & Hamilton, A. F. de C. (2008). Unbroken mirrors: challenging a theory of Autism. *Trends in Cognitive Sciences*, *12*(6), 225–229. http://doi.org/10.1016/j.tics.2008.03.005

Spoelstra, J., Schweighofer, N., & Arbib, M. A. (2000). Cerebellar learning of accurate predictive control for fast-reaching movements. *Biological Cybernetics*, *82*(4), 321–33. http://doi.org/10.1007/s004220050586

Tagamets, M.-A., & Horwitz, B. (1998). Integrating Electrophysiological and Anatomical Experimental Data to Create a Large-scale Model that Simulates a Delayed Match-to-sample Human Brain Imaging Study. *Cerebral Cortex*, *8*(4), 310–320. http://doi.org/10.1093/cercor/8.4.310

Tanaka, K. (1996). Inferotemporal Cortex and Object Vision. *Annual Review of Neuroscience*, *19*(1), 109–139. http://doi.org/10.1146/annurev.neuro.19.1.109

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. http://doi.org/10.1016/j.neunet.2004.05.007

Tani, J., Nishimoto, R., & Paine, R. W. (2008). Achieving "organic compositionality" through self-organization: Reviews on brain-inspired robotics experiments. *Neural Networks*, *21*(4), 584–603. http://doi.org/10.1016/j.neunet.2008.03.008

Thiel, C. M., Zilles, K., & Fink, G. R. (2004). Cerebral correlates of alerting, orienting and reorienting of visuospatial attention: an event-related fMRI study. *NeuroImage*, *21*(1), 318–28. http://doi.org/10.1016/j.neuroimage.2003.08.044

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., … Belliveau, J. W. (1995). Functional Analysis of Human MT and Related Visual Cortical Areas Using Magnetic Resonance Imaging. *The Journal of Neuroscience*, *15*(4), 3215–30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7722658

Triantafyllou, C., Hoge, R. D., Krueger, G., Wiggins, C. J., Potthast, A., Wiggins, G. C., & Wald, L. L. (2005). Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*, *26*(1), 243–250. http://doi.org/10.1016/j.neuroimage.2005.01.007

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I Know What You Are Doing: A Neurophysiological Study. *Neuron*, *31*(1), 155–165. http://doi.org/10.1016/S0896-6273(01)00337-3

Vaillancourt, D. E., Mayka, M. A., & Corcos, D. M. (2006). Intermittent visuomotor processing in the human cerebellum, parietal cortex, and premotor cortex. *Journal of Neurophysiology*, *95*(2), 922–931. http://doi.org/10.1152/jn.00718.2005

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. http://doi.org/10.1016/j.neunet.2005.07.014

Want, S. C., & Harris, P. L. (2001). Learning from Other People's Mistakes: Causal Understanding in Learning to Use a Tool. *Child Development*, *72*(2), 431–443. http://doi.org/10.1111/1467-8624.00288

Wheaton, L. A., Nolte, G., Bohlhalter, S., Fridman, E., & Hallett, M. (2005). Synchronization of parietal and premotor areas during preparation and execution of praxis hand movements. *Clinical Neurophysiology*, *116*(6), 1382–1390. http://doi.org/10.1016/j.clinph.2005.01.008

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*(4), 287–295. http://doi.org/10.1016/S0149-7634(01)00014-8

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. http://doi.org/10.1016/S1364-6613(97)01070-X

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 593–602. http://doi.org/10.1098/rstb.2002.1238

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. http://doi.org/10.1038/2245

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7-8), 1317–1329. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12662752

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347. http://doi.org/10.1016/S1364-6613(98)01221-2