2015 Special Issue

Exploiting the gain-modulation mechanism in parieto-motor neurons: Application to visuomotor transformations and embodied simulation

Sylvain Mahé, Raphaël Braud, Philippe Gaussier, Mathias Quoy, Alexandre Pitti *

Laboratoire ETIS - UMR CNRS 8051, Université de Cergy-Pontoise, Bat. St-Martin, 2, avenue Adolphe-Chauvin, F 95302 Cergy-Pontoise Cedex, France

HIGHLIGHTS

- We exploit the gain-field effect in parietal neurons for sensorimotor transformations.
- Construction of a body map is based on visuo-motor integration in a robotic arm.
- Error between real and estimated signals models the hidden spatial transformation.
- This feature of gain-fields neurons is used to solve the correspondence problem.
- Gain-field neurons learn external-point reference frames for tool-use and body change.

ARTICLE INFO

ABSTRACT

The so-called self–other correspondence problem in imitation demands to find the transformation that maps the motor dynamics of one partner to our own. This requires a general purpose sensorimotor mechanism that transforms an external fixation-point (partner's shoulder) reference frame to one's own body-centered reference frame. We propose that the mechanism of gain-modulation observed in parietal neurons may generally serve these types of transformations by binding the sensory signals across the modalities with radial basis functions (tensor products) on the one hand and by permitting the learning of contextual reference frames on the other hand. In a shoulder–elbow robotic experiment, gain-field neurons (GF) intertwine the visuo-motor variables so that their amplitude depends on them all. In situations of modification of the body-centered reference frame, the error detected in the visuo-motor mapping can serve then to learn the transformation between the robot's current sensorimotor space and the new one. These situations occur for instance when we turn the head on its axis (visual transformation), when we use a tool (body modification), or when we interact with a partner (embodied simulation). Our results defend the idea that the biologically-inspired mechanism of gain modulation found in parietal neurons can serve as a basic structure for achieving nonlinear mapping in spatial tasks as well as in cooperative and social functions.

1. Introduction

Over the last two decades, the studies of the post-parietal cortex (PPC) have permitted to understand better the neural mechanisms involved in the spatial representation of oneself body. What we have discovered is that our body representation is far more labile as we previously thought and that the brain fully exploits the perceptual ambiguity yielded by the senses to represent spatially not only our body limbs but also the nearby objects and by extension the persons around us. The so-called mirror neurons found by Rizzolatti and his colleagues exemplify the most this discovery as they respond to action and to observation (Rizzolatti, Fadiga, Fogassi, & Gallese, 1996; Rizzolatti, Fogassi, & Gallese, 2001). Mirror neurons appear to bear out the fundamental structure for achieving perceptual, cognitive and motor functions as well as cooperative and social functions (Fogassi et al., 2005; Keysers, 2004) although its mechanism is still poorly understood.

In this perspective, the studies by Iriki in the macaque monkeys are particularly interesting as they showed evidences of a dynamical readaptation of the body schema with respect to the ongoing situation (Iriki, Tanaka, & Iwamura, 1996). By simply manipulating

* Corresponding author. Tel.: +33 81761992602.
E-mail address: alexandre.pitti@u-cergy.fr (A. Pitti).

the visual feedback on a TV set that a monkey scrutinizes to guide its arm motion, Iriki showed how the parietal neurons were readjusting continuously the body image (here the hand) in accordance to the new reference frame (Okanoya, Tokimoto, Kumazawa, Hihara, & Iriki, 2008). The spatial transformations performed could be as complex and nonlinear as the combination of translation, rescaling and rotation. This result was also tested on tool-use where the spatial receptive fields of the parietal neurons associated to the hand extended to entail the tool (Goldenberg & Iriki, 2007; Maravita & Iriki, 2004). In terms of social cognition, this transformation mechanism is considered to take a central place in the process of understanding others as a mean to transform someone else visuo-motor perception into our own thus simulating their actions (Fogassi et al., 2005; Lewkowicz, Delevoye-Turrell, Bailly, Andry, & Gaussier, 2013; Meltzoff, 2007; Rizzolatti et al., 2001), see Fig. 1.

Thus, one may recognize that the neural mechanisms involved in spatial representations constitute a hard problem that requires at the same time non-linear transformations as well as rapid processing. On the one hand, the PPC is ideally placed for multimodal integration since it is one of the first cortical structure to receive the sensory signals coming from the different modalities (Andersen, 1997; Pouget & Snyder, 1997). On the other hand, its role to bind fastly the sensory signals is not trivial at all since each sensory signal is encoded differently and anchored to different body part or spatial reference frame; e.g., eye-, head-, shoulder- or hand-centered. One consequence is that patients with lesions of the parietal cortex present difficulty in spatial adjustment, coordination disorders and even spatial neglect (Keysers, 2004). Moreover, the spatial disorders also pervade in the social domain particularly in autism spectrum disorders with the importance of embodied self-rotation for visual and spatial perspective-taking (Pearson, Ropar, & Hamilton, 2013; Surtees, Apperly, & Samson, 2013). These studies have revealed a lack of multimodal integration and the disability to put in perspective the spatial location of objects and persons relative to our body.

Considering the mechanisms it may involve, the discovery of reach cells and postural cells for particular orientation of the hand associated to the current context or motor plan have permitted to discriminate further the functional organization at the network level (Blohm & Crawford, 2009; Bremner & Andersen, 2012; McGuire & Sabes, 2009). Andersen and colleagues (Andersen, Essick, & Siegel, 1985; Andersen & Mountcastle, 1983) first discovered neurons firing for a specific eye saccade motor command, *modulated* by the position of the eye relative to the head. That is, this result demonstrates that (1) these neurons are bimodal neurons as they encode two information at once and that (2) their amplitude level is an informative quantity that can be modeled. Furthermore, the gain-modulation effect observed for this behavior does not correspond to a summing integration as it would be for integrate-and-fire neurons. Instead, a more correct mathematical model of the parietal neurons' response would be a multiplicative integration between the incoming sensory signals, which can be approximated as a nonlinear basis function (Pouget & Snyder, 1997).

The striking advantage of a gain-field representation of the signal is that a basis function representation may approximate any desired mapping, as it is the case for the Fourier series or the wavelet decomposition. Therefore, this kind of representation meets the requirements for a local-to-allocentric transformation because multiple reference frames can be derived from the same population of neurons allowing to use intrinsic as well as extrinsic reference frames (Bremner & Andersen, 2012; McGuire & Sabes, 2009). For instance, Shadmehr and Wise proposed that the gain-field neurons compute a fixation-centered frame by subtracting the vector between the gaze location and the hand position to
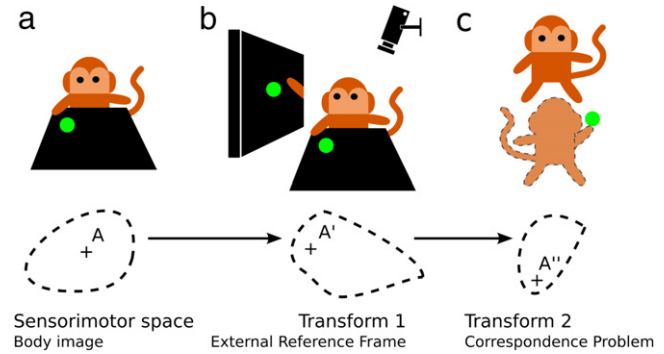
Fig. 1. The problem of the frame of reference and of sensorimotor transformation. (a) Grasping is a complicated task as it requires to learn the visuomotor space in an ego-egocentric reference frame. (b) When we observed our own action in a TV set, we change reference frame, which requires to transform the spatial information of the visual coordinates with respect into the hand coordinates. (c) The same situation occurs when we observe someone else's actions and try to imitate them; this transformation is called the correspondence problem. The point A in visual space corresponds to the point A' in the new reference frame after transformation and to the point A'' after a different transformation.

derive the hand toward the target in eye-centered frame (Bremner & Andersen, 2012; Shadmehr & Wise, 2005). Following this, different robotic experiments have been conceived using the linear combination of basis functions for sensorimotor transformations (Chinellato, Antonelli, Grzyb, & del Pobil, 2011; Halgand, Soueres, Trotter, Celebrini, & Jouffrais, 2010; Hoffmann et al., 2010).

In previous works, we demonstrated how the mechanism of gain-field modulation can be applied for integrating audio–visual signals and proprioceptive feedback in a head–neck–eyes robotic device as it is for some parietal neurons (Pitti, Blanchard, Cardinaux, & Gaussier, 2012). In our studies, the gain-field neurons were successfully used to remap the location of one sound signal (in head-centered reference frame) into retina coordinates (in eye-centered reference frame). The gain-field based model enabled the system to increase the accuracy of a visual stimulus (i.e., the position of the mouth when a person speaks) by using the supplementary sound signal to estimate more precisely the spatial location of the mouth–voice stimulus.

In this paper, we consider to employ again the gain modulation mechanism but this time toward a fixation-point reference frame, external to the body (Shadmehr & Wise, 2005). We propose that the same neural architecture may permit to derive cognitive functions involved in fixation-point reference frame (such as tool-use) as well as social functions involved in perspective-taking tasks such as joint attention and imitation. To this end, we first perform a rapid learning of the visuomotor associations in a gain-field architecture with a shoulder–elbow-like robotic arm and a camera. Once its body schema is learned, a second gain-field module is added for situations of visuomotor mismatch and novelty. This second module is used to encode the new sensorimotor *task set* (Pitti, Braud, Mahé, Quoy, & Gaussier, 2013a; Pitti, Mori, Kouzuma, & Kuniyoshi, 2009) corresponding to the visuomotor distortions induced by novel fixation-centered tasks such as in front of a mirror, during tool-use or during the control of an avatar in a video game (self-observation in a TV set). In the social domain, the idea is then to retrieve the hidden visuomotor transformation responsible for the mismatch between the robot arm's own motion and what it currently sees (e.g., a person moving his/her hand aside). The hidden visuomotor transformation permits to reduce the resulting spatial error in the visual field and in the motor domain, which is associated to the so-called correspondence problem (Brass & Heyes, 2005; Heyes, 2001) and to motor imaginary (Kosslyn, Ganis, & Thompson, 2001).

2. Methods

2.1. A formal model of gain fields

Our architecture implements multiplicative neurons, called gain-field neurons, that multiply unit by unit the value of two or more incoming neural populations, see Fig. 2. Its organization is similar to radial basis function (RBF) as it transforms the incoming signals into a representation of basis functions, a *functional* space, that could be exploited to simultaneously represent stimuli in various reference frames (Pouget & Snyder, 1997; Salinas & Thier, 2000). The multiplication between afferent sensory signals from two different modalities (M_1, M_2) is the element-wise multiplication between two probability distributions X_{m_1} and X_{m_2}, two vectors of dimensions $\{\mathbb{M}_1$ and $\mathbb{M}_2\}$ respectively, with $\{m_1, m_2 \in \mathbb{M}_1, \mathbb{M}_2\}$. The featured matrix is the signal activity X_m, $m \in \mathbb{M}_1 \times \mathbb{M}_2$ that gain-field neurons learn the activity. The equation for nth gain-field neuron X_n^{GF}, $n \in N$, with synaptic weights w_i, $i \in \mathbb{M}_1, \mathbb{M}_2$ is:

$$X_m = X_{m_1} \times X_{m_2} \tag{1}$$

$$X_n^{GF} = \sum_{i=0}^{\mathbb{M}_1 \times \mathbb{M}_2} w_i X_i. \tag{2}$$

Hence, neurons X^{GF} have the same equation as perceptrons without bias. The key idea here is that X^{GF} map encodes a particular combination of the two values, the amplitude of the gain-field neurons encode one modality *conditionally* to another in a Bayesian fashion, in a lower dimension (Braun, Aertsen, Wolpert, & Mehring, 2009; Braun, Mehring, & Wolpert, 2010). We exploit this feature to model the parietal circuits from different sensory signals as a linear combination of gain-fields neurons f so that we have $f(X_n^{GF}) = \sum_{n=0}^{N} \omega_n X_n^{GF}$, for $n \in N$ gain-field neurons, and ω the pondering weights. In a sense, the linear combination of perceptron-like neurons corresponds to the neural architecture of RBFs.

We explain hereinafter how the gain-fields neurons learn the associations between various modalities, see Fig. 2. Once a first mapping is done – say between variables X and Y in Fig. 2 – it is possible to chain the gain-field maps so that the activity of the latters (e.g., to encode the new modality θ_1) depends on the activity of the formers (i.e., the first modalities X and Y). As an example, we give the equations for the second map X_n^{GF2} constructed from the activity of a first map $f(X_n^{GF1})$ and the new modality X_{m_3}, a vector of dimension $\mathbb{M}_3$:

$$X_m' = X_{m_3} \times f(X_n^{GF1}) \tag{3}$$

$$X_n^{GF2} = \sum_{i=0}^{\mathbb{M}_3 \times N} w_i X_i'. \tag{4}$$

The new gain fields neurons X_n^{GF2} depend on X_{m_3} and $f(X_n^{GF1})$ and encode then a relational map obtained from the activity of all the preceding modalities, X_{m_1} and X_{m_2}, with the new one, X_{m_3}. We make a point that X^{GF1} represents essentially the multiplication of the two input modality values, which do not uniquely represent one combination of the two values. However, from a biologically viewpoint, the observation of single gain-field neuron shows a similar behavior in parietal neurons (Salinas & Thier, 2000). While one X^{GF} is not enough to disambiguate certain cases, the linear combination of multiple X^{GF1}, as X_n^{GF2} do, can do much for generalization, as it is for RBFs and perceptrons. Thus, the GF neurons each represent a specific overall activity pattern of the X^{GF} map which encodes a pair-wise combination of the two values. By doing so, the system retranscribes the relative information of the occurrence of one modality with respect to the others.
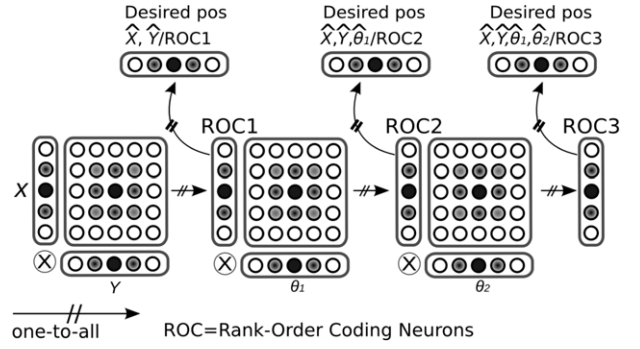
Fig. 2. Neural architecture for gain-field neurons. Gain-field neurons can bind the activity of one, two, three or four inputs by reusing gain-field neurons resulting from the multiplication of two modalities as one of the two modalities of another gain-field map. These neurons can serve as basis functions from which desired representations can be mapped. (ROC) stands for Rank-Order Coding neurons. The sensory signals I can be estimated then from the linear combination of ROC neurons as it is with perceptron neurons. The desired output, $\hat{I}$/ROC, means then the estimation of signal I by ROC.

This multimodal process is similar to the chaining of Bayesian conditional rules between multiple variables (Deneve & Pouget, 2004). As the number of dimension augments, a linear combination of GF neurons can permit to represent well the input data for an appropriate number of GF neurons, if the input space is sparse enough. In this case, we should not see a resolution loss after the learning stage.

2.2. The Rank-Order Coding algorithm

In order to learn the sensorimotor mapping between input and output signals, we implement the hebbian-like learning algorithm proposed by Thorpe and colleagues (Thorpe, Delorme, & Van Rullen, 2001; Van Rullen, Gautrais, Delorme, & Thorpe, 1998) called the Rank-Order Coding (ROC) algorithm that we have used already in previous researches (Pitti et al., 2012, 2013a). The ROC algorithm has been proposed as a discrete and faster model of the derivative integrate-and-fire neuron (Van Rullen & Thorpe, 2002). ROC neurons are sensitive to the sequential order of the incoming signals; that is, its *rank code* and the distance similarity to this code is transformed into an amplitude value. A scalar product between the input's rank code with the synaptic weights furnishes then a distance measure and the activity level of the neuron. More precisely, the ordinal rank code can be obtained by sorting the signals' vector relative to their amplitude levels or to their temporal order. Due to this feature to encode the rank, ROC neurons are more robust to amplitude noise than winner-takes-all (WTA) (Thorpe et al., 2001).

The neurons' output X^{GF} is computed by multiplying, not directly the amplitude values of the sensory signal vector I, but the inverse of its rank order $rank(I)$ by the synaptic weights w; $w \in [0, 1]$. For an input vector signal of dimension M and for a population of N GF neurons X, we replace Eq. (2) by:

$$X_n^{GF} = \sum_{m \in M} \frac{1}{rank(I_m)} w_{n,m}^{GF}. \tag{5}$$

The updating rule of the neurons' weights is similar to the WTA learning algorithm of Kohonen's self-organizing maps (Kohonen, 1982). For the best neuron $s \in N$ and for all afferent signals $m \in M$, we have:

$$\Delta w_{s,m}^{GF} = \alpha \left(\frac{1}{rank(I_m)} - w_{s,m}^{GF} \right), \tag{6}$$

where α is the learning rate; we set $\alpha = 0.01$. We make a note that the synaptic weights follow a power-scale density distribution that
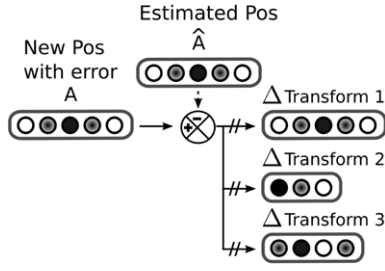
Fig. 3. Transformation mechanism with an error signal. We can compute an error signal ΔA between a variable A and its estimation $\hat{A}$ done by the neural network in Fig. 2. This error can be learned by different maps with respect to the context it corresponds. The error map can serve then to learn the transformation to pass from the point A to the point $\hat{A}$.

makes the Rank-Order Coding neurons similar to basis functions. This attribute permits to use them as receptive fields so that the more distant the input signal is to the receptive field, the lower is its activity level.

2.3. Learning sensorimotor transformation

Once the gain-field neurons have learned the visuomotor rules from their input space, they can be exploited for comparing any situation of visuomotor mismatches during physical perceptual changes as well as during social ambiguities. For example, a comparator model as plotted in Fig. 3 that computes the difference between a real signal value A with an estimated one $\hat{A}$ from the network can serve to retrieve back the possible transformations responsible for the sensorimotor errors.
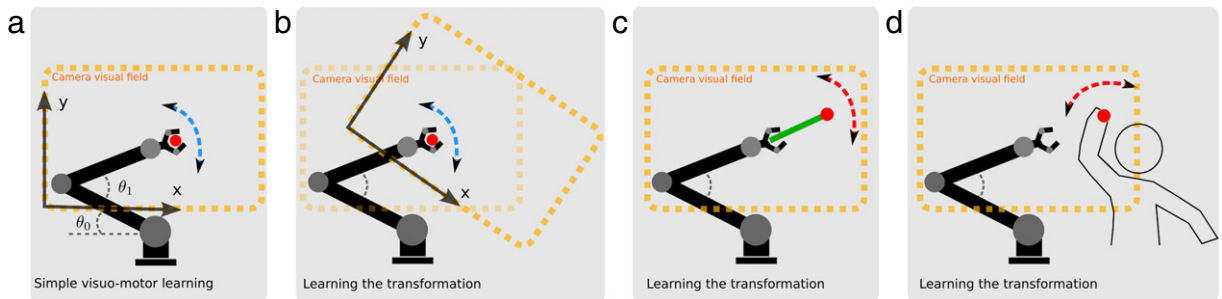
In that sense, the GF neurons used in our model are similar to radial basis functions (RBF) used in image processing for morphing and registration problems where the linear combination of RBF can serve to deform the input space to a particular a priori model. Let us consider the point A and the new point $\hat{A}$ estimated by the GF network with:

$$f(A) = \sum_{i=1}^{n} w_i X_i^{GF}(A)$$
$$f(\hat{A}) = \sum_{i=1}^{n} w_i X_i^{GF}(\hat{A}).$$
(7)

Let us express $f(\hat{A})$ as a function of $f(A)$:

$$f(\hat{A}) = \sum_{i=1}^{n} w_i(X_i^{GF}(A) + (X_i^{GF}(\hat{A}) - X_i^{GF}(A)))$$

$$f(\hat{A}) = \sum_{i=1}^{n} w_i X_i^{GF}(A) + \sum_{i=1}^{n} w_i(X_i^{GF}(\hat{A}) - X_i^{GF}(A))$$

$$f(\hat{A}) = f(A) + \sum_{i=1}^{n} w_i(X_i^{GF}(\hat{A}) - X_i^{GF}(A)).$$
(8)

This means that the closer $\hat{A}$ is to A, the closer $f(\hat{A})$ will be to $f(A)$, in other words, the closer a point $\hat{A}$ is from a point A used for the training of the GF neurons, the closer the associated transformation will be to the example transformation.

The second term corresponds to the transformation function that minimizes the distance between $X_i^{GF}(A)$ and $X_i^{GF}(\hat{A})$. One perceptron can estimate this function within the range around the point A by identifying the weights that minimize the correspondence problem, Although a population of perceptrons is more satisfactory since it can cover the whole space as well as different transformation functions with respect to the current situation.
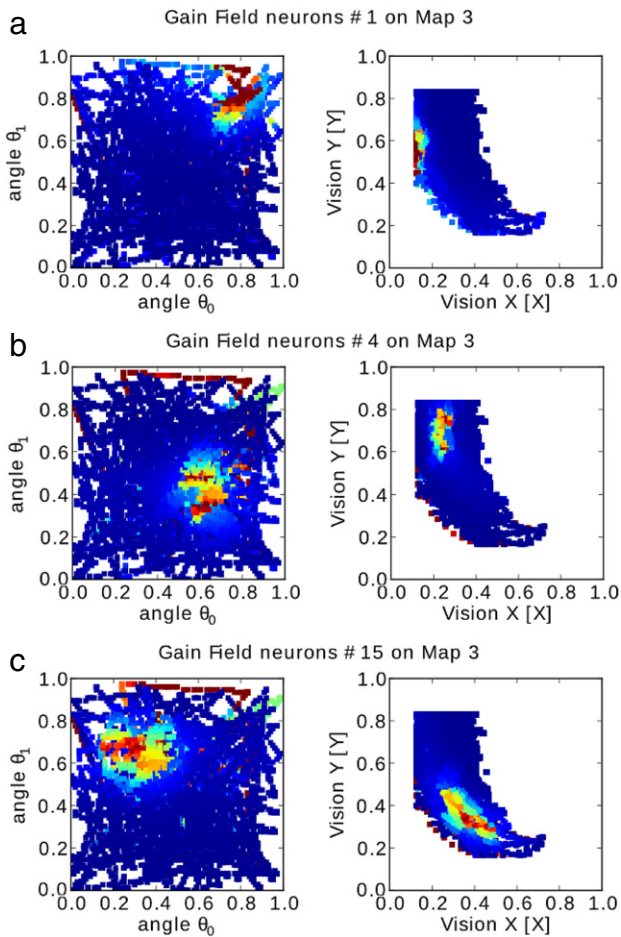
3. Experimental setup

In our experiments, we use the Kinova robotic arm with 7 degrees of freedom and a fixed camera to observe it, a graphic of the setup is shown in Fig. 4(a).

For the sake of simplicity, we limited the arm to two Degrees-of-Freedom (DoF) in the visual plan of the camera, but the architecture can be easily extended to more DoFs by chaining the GF maps. The two motor joints correspond respectively to the shoulder and elbow-like joints, θ_0 and θ_1; with $\theta_0 \in [0°, 200°]$ and with $\theta_1 \in [0°, 100°]$. Each motor angle is translated into a discretized vector of 22 bins with a Gaussian curve centered on the current motor angle. A color-based vision system focusing on the red color provides the hand's coordinates (x, y) on the retina reference frame. And similar with the motor angles, the visual coordinates are translated into two discretized vectors of 22 bins with a Gaussian curve centered on the current position. Using the gain-field architecture as in Fig. 2, the system is composed of three maps, where each ROC group is composed of 22 neurons in order to have a regular network. The last ROC group learns to associate then the motor angles (θ_0, θ_1) with the coordinates of the hand in the (x, y) axis. The output of the gain-fields network will now give an RBF based representation of the learned (x, y) coordinates of the hand.

To take advantage of the gain-fields, an output layer of perceptron neurons will combine their amplitude level toward the estimation of the hand's coordinates (x, y) in the visual field and of the motor angles; resp. $(\hat{x}, \hat{y})$ and $(\hat{\theta}_0, \hat{\theta}_1)$. This will correspond to the integration of the four inputs through three chained maps as in Fig. 2; see also Section 2.1.

Fig. 4. Experimental setups and sensorimotor transformations. (a) The first situation corresponds to the learning of the visuomotor correspondence between the motor variables (two degrees of freedom) and the (X, Y) position in the camera visual field. Although this task is nonlinear, a slight change of the reference frame does not require to relearn all the visuomotor links but instead, the transform function responsible for the global error. These changes can be due to a visual transformation as for a camera rotation (b) or to the body changes as during tool-use (c) or to the embodied simulation of one partner as in (d); the so-called correspondence problem. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Activity of Gain Fields neurons. Neurons 1, 4 and 15 resp. (a)–(c) encode different sensorimotor locations in (θ_0, θ_1) and (X, Y) as it would be for radial basis functions. Each variable is normalized within the interval [0, 1]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Amplitude dynamics of one Gain-field neuron with respect to the motor variable θ_0 on the X axis. It is noteworthy that the two variables are linked together with respect to the amplitude. The X centers slightly shift also with respect to the variable θ_0 as it is seen in biological neurons. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
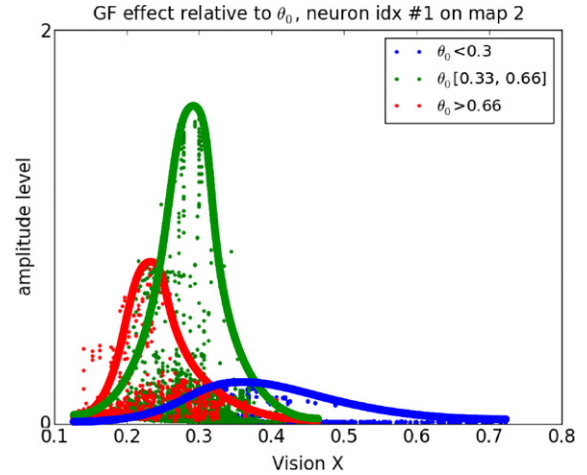
Then, in the second part of the experiment, we will use this output layer in order to learn the transformations produced in situations of perceptual ambiguities; i.e., camera-shift, tool-use and social interaction, see Section 2.3 and Fig. 4(b)–(d). A third layer will be added that will learn the difference between the estimated values of the hand coordinates and the current ones (after the transformation is applied to the system); c.f., Eq. (8). The output of the third layer of perceptron neurons provides the transformation for any given point that will cancel the transformation applied to the system and thus making the previously learned action still relevant.

4. Results

4.1. Generalization from visuomotor association

We first let the network learn the visuomotor coordination with the gain-field neurons and we check then how the learning operates toward the spatial location of the arm in the eye-field and in the peripersonal space.

At the initial stage, we perform a motor babbling in a randomized fashion so that the ROC neurons will map uniformly the visuomotor space. After the learning period, which corresponded to 12 min duration, 12,000 samples with the exploration of a new arm configuration each 0.05 s, the ROC neurons self-organized themselves and reproduced the gain-field behavior of PPC neurons.

For instance, we plot in Fig. 5 the activity level of three different ROC neurons from the third map with respect to the variables pairs (θ_0, θ_1) and (x, y). The color code indicates the activity value for each neuron and their respective receptive fields in the motor domain and in the visual domain. We make a note that the motor and visual receptive fields are not separated from each other, but the activity level of the gain-fields neurons is related altogether to the four variables: in Fig. 6 for instance, the amplitude level for one GF neuron varies with respect to the motor angle θ_0 for a given range of values on the X axis; the color code indicates three different intervals of the motor angle θ_0. We can see that its receptive field is not strictly centered on a particular value on the X axis but shifts with respect to the proprioceptive feedback; a phenomenon also observed in the biological PPC neurons (Salinas & Thier, 2000).

One consequence for gain-fields neurons to simulate basis functions is that the linear combination can serve to estimate the current visuo-motor state in a specific reference frame. Gain-modulated neurons in PPC have been found to generate depth-dependent activity (Blohm & Crawford, 2012) as well as hand-centered maps (Galeazzi et al., 2013) but because our setup does not have vergence and tactile information, we will not focus on these properties. Instead, we propose to use the gain-field neurons to estimate each incoming signal from the fusion of all, in their respective reference frame.

A population of perceptron neurons is used to adapt its weights correctly to the appropriate linear combination of gain-fields in order to estimate the (x, y) variables, $(\widehat{x}, \widehat{y})$; see Fig. 7(a). The blue line corresponds to the real visual location on the Y axis whereas the green line and the red line are the linear combinations learned by the perceptron neurons respectively from the first ROC map $(\widehat{y}/\text{ROC1})$ and from the third ROC map, $(\widehat{y}/\text{ROC3})$; the first map is the one that receives information from the (x, y) position solely whereas the third map combines the information coming from the four variables. In this plot we observe a slight variance and bias in the estimation process for the two output networks, more pronounced for the second system, which integrates the motor signals to the visual ones. The histograms plotted in Fig. 7(b) display the normalized spatial error for the two neural populations. The two histograms correspond to two power-law curves but the variation in their shape exemplifies how the two neural networks generalize also differently. The larger error variance in the second perceptron map indicates that the motor variables (θ_0, θ_1) have a defavorable influence in the estimation of the visual variables.
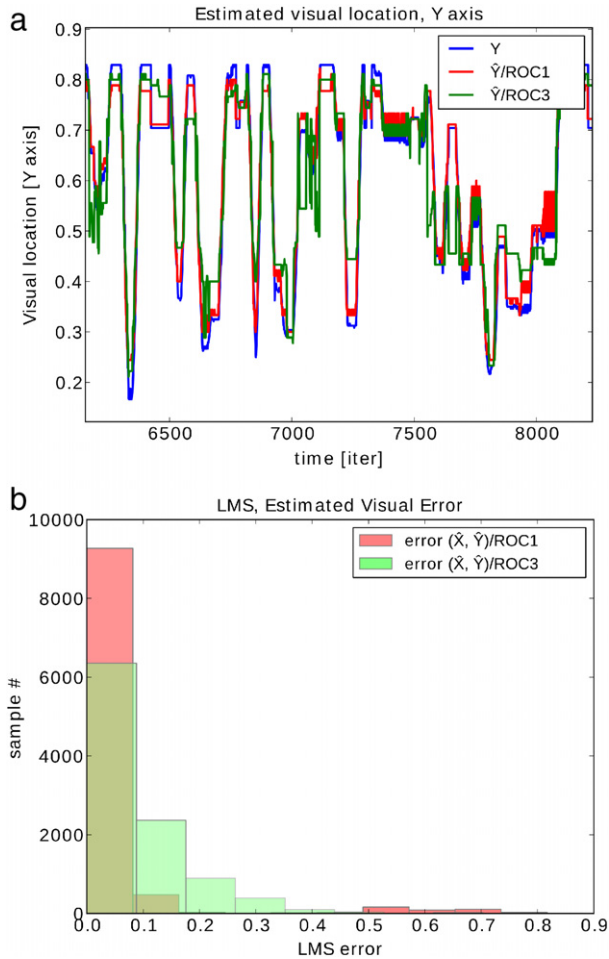
Fig. 7. Neural estimation on the visual field. (a) The blue line corresponds to the location of the target to the *Y* axis whereas the red line and the green line correspond respectively to the estimated location of the two output neural networks respectively calculated from the ROC neurons from the first map (visual inputs only) and from the ROC neurons from the third map (visual and motor inputs). (b) Histogram of the visual error for the two neural networks, the variance of the error distribution is larger for the network estimated from the visual and motor inputs (green) than from the sole visual inputs (red), although some big errors occur above 0.5: the motor inputs have an influence on the estimation of the spatial location of the visual targets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Neural extension to visuomotor transformation

To estimate the generalization properties of our system for visuomotor transformation, we first perform our experiments on three basic visual transformations as done by Iriki on a macaque in which the primate showed rapid adaptation to new visual position of its limbs with changes operated on the screen (Iriki et al., 1996) or on the body (Maravita & Iriki, 2004; Okanoya et al., 2008). The first transformation consists on the small rotation of the camera in the plane of the robotic arm and the second transformation corresponds to a body extension of the robot hand with a tool (tool-use), the third transformation consists on a translational shift of the camera with respect to the arm plane of motion; see resp. Fig. 4(b) and (c).

We present the results for the estimated transformations of three population outputs for each case presented above in Fig. 8(a), (b) and (c) respectively for rotation, translation and tool-use. Each perceptron neuron of the output layer estimates the $(\Delta x, \Delta y)$ values learned from the current values received (x, y) and the expected ones (x', y') from the GF maps.
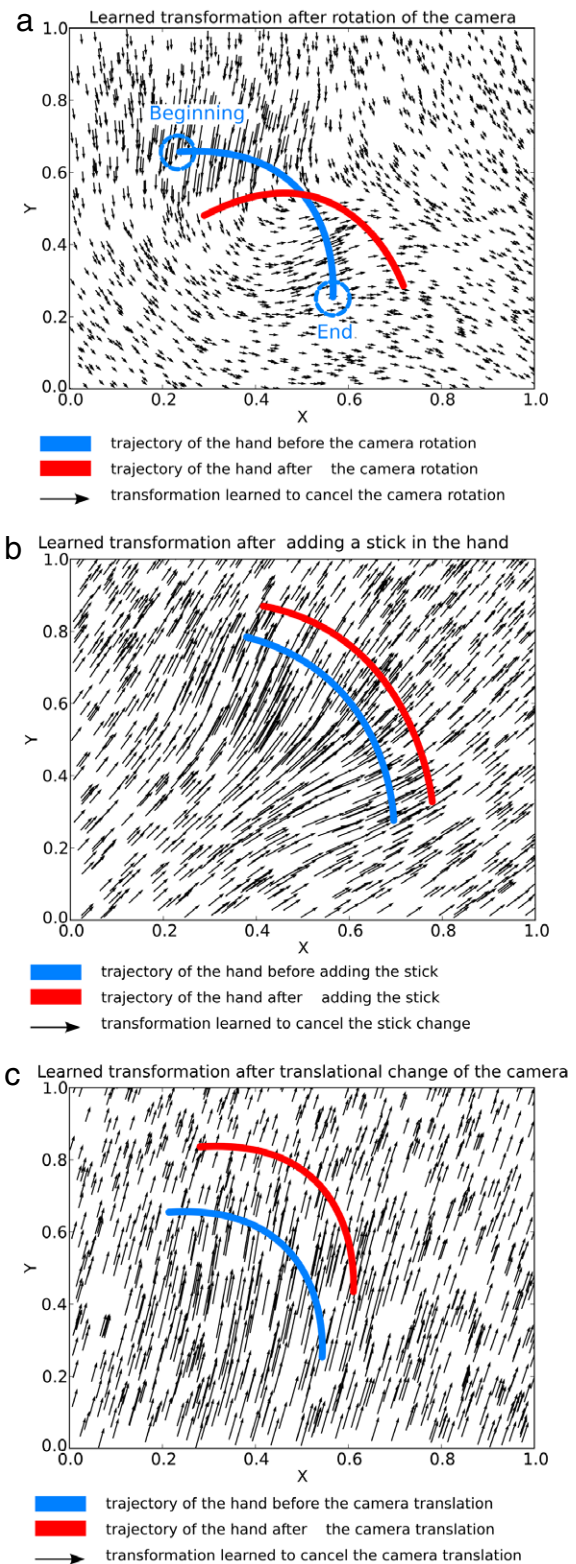
Fig. 8. Rotation, Tool-use and Shift Transformations. The arrows correspond to the estimated local transformations, the blue line corresponds to the original trajectory of the robot hand before any transformation and the red line corresponds to its trajectory after transformation. (a) Learned transformation after the rotation of the camera. (b) Learned transformation after adding a stick to the robot hand. (c) Learned transformation after the camera shifting. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to visualize the transformations that link the two sensorimotor spaces, we plot the arrows that link each point (x, y) with its associated ones $(x + \Delta x, y + \Delta y)$ after transformation. In Fig. 8, we plot in blue the trajectory of the hand before transformation and in red the trajectory of the hand after transformation. We make a note that the length of the arrows do not correspond to their actual size, they are reduced in order to be visualized better in the figure. Fig. 8(a) displays the transformation for the camera rotation. The blue line corresponds to the position of the hand before transformation and the red line corresponds to its position after transformation. As it can be seen, the arrows within the center, where the robot moves, show a correct approximation of the rotation as the generalization is made on the close neighborhood of the learned examples. However, past to a certain distance to the originally learned examples – and outside the center area – we can see some local discrepancies as the orientation of the arrows do not perfectly match the direction of the rotation anymore, although it is a global transformation.

The second transformation in Fig. 8(b) is the addition of a small stick at the end of the arm that effectively scales the arm up. During the second step of the experiment, the vision system would in this case recognize the position of the end of the stick instead of the position of the hand. Once again the transformation is successfully learned, and the generalization applies to a much larger part of the parameter space, which is probably due to the simplicity of the transformation. With respect to rotation, this transformation corresponds to a local transform, not applied on the whole image.

The last transformation performed is a simple vertical shift of the camera; see Fig. 8(c). Once again the transformation is successfully learned, and the generalization is applying to almost the whole parameter space, which differs from the previous experiment.

What is interesting to notice on the three presented results in Fig. 8 is the capacity of the system to generalize at least to a certain extent; see Fig. 9 for a quantitative measure of the root mean square error of the estimated location of the visual target. For each transformation, all the learning input (the conditional signal of the perceptron neurons) were gain-field representations of a point on the red line, while the associated example (the unconditional link of the perceptron neurons) was on the blue line. This coverage of the parameter space is far from complete, however the generalization of the transformation to the rest of the parameter space is rather correct, at least in terms of direction. If we look closer, for any given point, the generalized transformation corresponds to a fraction of the closest learned example. If a point is in between two learned examples, then the transformation is an interpolation of these two example transformations.

4.3. Motor imagery & perspective-taking in social context

The previous experiments were resolved with the estimation of the spatial error in the visual field. In comparison to body-centered sensorimotor transformation, the motor imaginary and perspective-taking tasks in the social domain require to observe the actions of one partner *and* to simulate his/her movements through our own motor system.

We investigate therefore how it is possible to estimate the motor configuration of someone else just from the visual input and to imitate one's arm posture with respect to what is seen. This situation corresponds to the so-called correspondence problem in social interaction for representing someone else's body posture (Brass & Heyes, 2005); see Figs. 1(c) and 4(d). Since there is no specific correspondence associations during role-taking, the new reference frame has to be computed from the estimated visual error from which the new motor action is calculated.
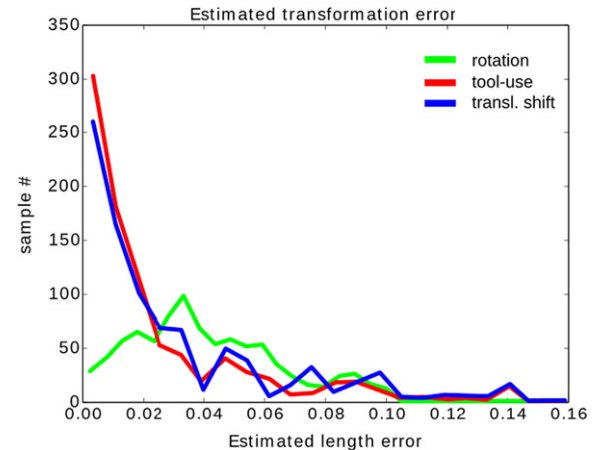

Fig. 9. Root mean square error computed from the position of the visual target and the estimated one for rotation (green), tool-use (red) and shift transformations (blue); resp. (a)–(c) in Fig. 8. The linear combination of gain-field neurons permit to reduce the error to a small range even for three different nonlinear transformations.

The experiment is as follows. The robot arm performs first a dynamic exploration of its arm visual location with a red toy in its hand and once it has learned its sensorimotor rules, a participant comes to interact with it by exploring his/her own peripersonal space with the same toy; see Fig. 10(a) and (b). We display in Fig. 10(c) with red arrows the visual transform found by the output neurons, which corresponds to a combination of a translational shift in the X axis with a rotational effect centered on an external fixation point, roughly situated at the shoulder's location of the student. This situation combines the three experiments done in Section 4.2 and Fig. 8.

Once the visual transform is defined, we analyze how the motor dynamics are estimated. We plot for this aim the robot personal space and the person's sensorimotor space in Fig. 11(a) respectively in blue and in red lines, where the arrows correspond to the visual transformation. We note that the two locations are different in space, therefore the neural architecture can differentiate its own robotic arm from one person's arm.

Considering the motor estimation, the top chart of Fig. 11(b) displays the transition in the visual signal in the X and Y axes from the period when the robot waves its hand and stops ($t < 500$ iterations) to the one when the student starts to wave her hand ($t > 500$ iterations). The middle chart exposes the motor dynamics (θ_0, θ_1) and the bottom chart displays the activity of the perceptron neurons involved in the motor transform for estimating the motor variables $\hat{\theta}_0$ and $\hat{\theta}_1$. We can observe from the graph that the output variables $\hat{\theta}_0$ and $\hat{\theta}_1$ estimated by the neural system in the motor domain (bottom chart) are replicating the variations of the visual signal (top chart), which are mostly on the Y axis and for the motor variable θ_0. The dynamics of the two variables are mostly in phase with the visual dynamics whereas the amplitude range is different. A plot of the renormalized estimated motor configuration is shown in Fig. 11(c) with respect to the visual stimulus. The neural system has mimicked the motor pattern of its partner and estimated its respective transform function.

5. Discussion

The Mirror Neurons System (MNS) shapes the sight information to the motor system when grasping an object. Its features extend to action recognition in the premotor cortex when it is the agent who is the observer so that the same cortical sites are activated during execution and observation. This system suggests that object-directed actions, communication and body movement are intertwined.
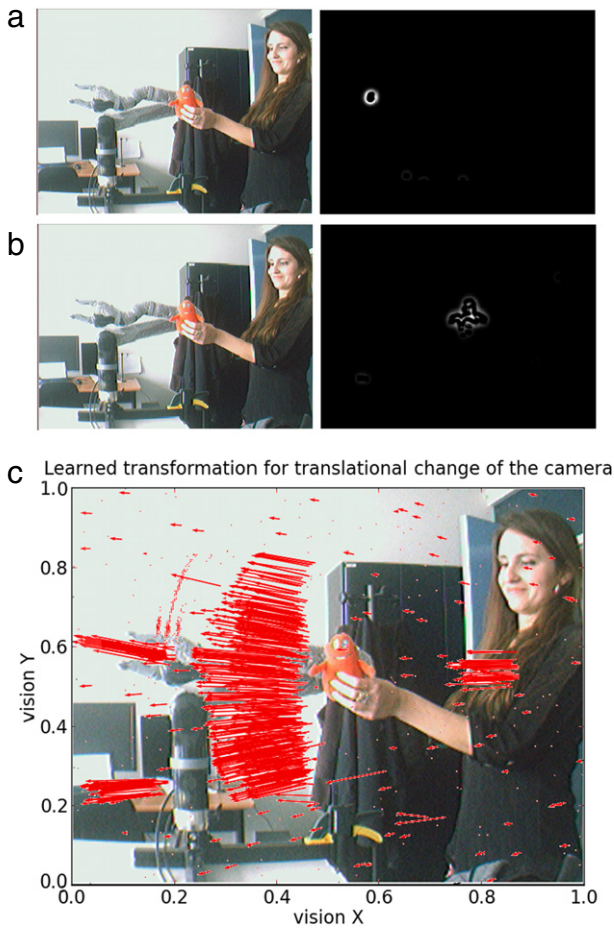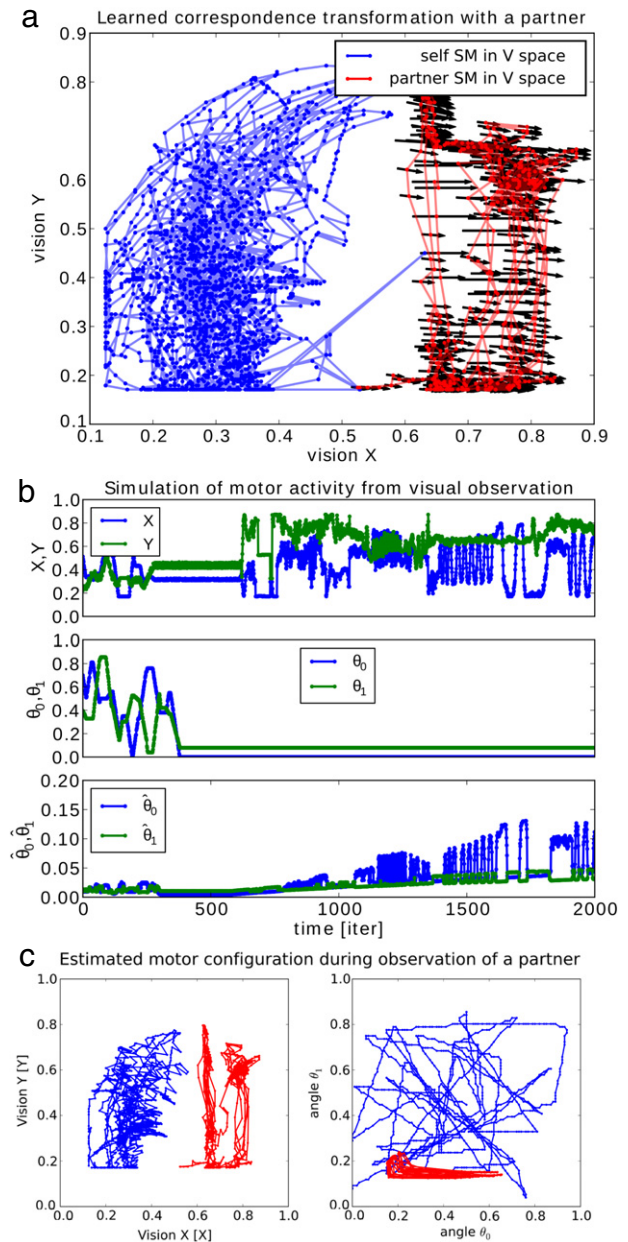
Fig. 10. Visual transformation and motor imagery during interaction with a partner. (a) and (b) Red toy handled by the robot and then handled by the student. This experiment reproduces the so-called correspondence problem posed in social interactions where one has to simulate the motor activity of one person into his/her own visuomotor system in order to imitate her or to recognize her intentions. (c) Within our framework, this simulation can be modeled as a visuomotor transformation by which the robot can learn the visual error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. Visual transformation and motor imagery. (a) The arrows display the estimated visual transformation between the robot peri-personal space (blue) and the partner space (red). This function can serve for self–other discrimination purpose as well as the estimation of the motor activity of the partner. In (b) and (c), the plots of the (X, Y) visual signals with the estimated motor configuration $(\hat{\theta}_0, \hat{\theta}_1)$ (normalized in (c)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

What is so special in the mechanism called embodied simulation is that we reuse our mental state (Gallese & Sinigaglia, 2011; Jeannerod, 2001). Embodied simulation should not be seen as a passive process by which mental states are "replayed", instead, it should be seen as an active process by which a transformation function is discovered online, mapping what it is seen to our own body dynamics. We propose that the processing involved in motor simulation in social cognition is a byproduct of the one involved in spatial transformation. We advance also that the gain-modulation effect in parietal neurons is part of the mechanisms responsible for it.

In previous works, we proposed a development scenario for the emergence of the MNS in two stages (Pitti & Kuniyoshi, 2012; Pitti, Mori, Yamada, & Kuniyoshi, 2010). We hypothesized that an automatic mimicry system is maturing at the fetal stage possibly in the superior colliculus (SC) forming the ground for the cortical circuits to organize the MNS (Pitti, Kuniyoshi, Quoy, & Gaussier, 2013b). At a first stage, we proposed that the topological alignment in SC across the modalities enables the automatic social responses of the newborns like facial preference and facial mimicry. This hypothesis has been also suggested by Nagy and Molnar (2004) and Neil, Chee-Ruiter, Scheier, Lewkowicz, and Shimojo (2006) who emphasized the central place that occupies the SC for fusioning the senses with respect to other brain regions not yet matured.

At a second stage, we propose then that a more complex spatial representation of the body in an allocentric metrics is emerging in the cortex as suggested in Bremner, Holmes, and Spence (2008), Del Giudice, Manera, and Keysers (2008) and Pitti et al. (2010). This body mapping would be based on the GF mechanisms of sensorimotor transformation responsible for body-centered reference frames (Andersen, 1997; Blohm, Khan, & Crawford, 2008; Salinas & Sejnowski, 2001) but also from the sensorimotor integration from the different modalities. As proposed by Shadmehr, such mechanism would permit to have far more plastic representations with fixation-point reference frames external to the body (Shadmehr & Wise, 2005). This system would permit to handle visuomotor changes such as during self-observation in a TV set or

during tool-use, but also for solving the correspondence problem in inter-personal interactions.

Considering the construction of the inter-subjective mind, the PPC has been well-acknowledged for its contribution to agency (Schwabe & Blanke, 2007; Tsakiris, Prabhu, & Haggard, 2006), intention (Andersen & Buneo, 2002; Andersen & Cui, 2009) and motor simulation (Kosslyn et al., 2001; Kosslyn, Ganis, & Thompson, 2006). Blanke – among other researchers – proposes that the sense of agency is related to the level of anticipation across the sensorimotor signals. Hiraki and colleagues suggest further that the synchrony among the sensorimotor signals activates certain parietal regions while its disrupting activates others (Miyazaki & Hiraki, 2006; Shimada & Hiraki, 2006; Shimada, Hiraki, & Oda, 2005) (e.g., with time lags). More precisely, the right inferior parietal cortex is consistently activated in conditions involving imitation (Decety & Chaminade, 2003; Meltzoff & Decety, 2003; Ruby & Decety, 2001). We suggest that this region is associated to self–other simulation and encodes the transformations necessary for extending the body representation.

If we consider now this problem from a robotic perspective, Nagai pointed out that no computational models that aim to address imitation nor the architecture of MNS took self–other discrimination into account (Nagai, Kawai, & Asada, 2011). For instance, these works take advantage of self-observation and perceptual ambiguity for the emergence of imitation (Andry, Gaussier, & Nadel, 2002; Andry, Gaussier, Nadel, & Hirsbrunner, 2004; Kuniyoshi, Yorozu, Inaba, & Inoue, 2003), but do not disambiguate the dynamics corresponding to oneself action and those of others (De Rengervé, Boucenna, Andry, & Gaussier, 2010). Among the few exceptions, the works done by Fuke, Ogino, and Asada (2009) and Nagai et al. (2011) propose to discriminate the motor space with respect to observed actions (self ones or others'). We can cite notwithstanding the hebbian solution proposed first by Keysers (2004) based on contingency detection at the neural level (i.e., the biologically-inspired mechanism of spike timing-dependent plasticity) for categorizing the sensorimotor signals in situation of interaction with others (Pitti et al., 2009). In comparison with these works, the present one emphasizes the learning of a transformation function that maps the robot visuomotor space (the egocentric reference frame) to the new one (the allocentric reference frame). We suggest that this mechanism of sensorimotor transformation based on gain-field neurons is general enough to serve for the mapping of spatial transformation during physical as well as social interactions.

Deneve and colleagues described in detail the advantages and disadvantages of representing multi-modalities using radial basis functions (tensor products). The main advantage of basis functions is notably to reduce the computation of nonlinear functions into linear ones (Deneve & Pouget, 2003) especially for retinal images, which are highly nonlinear, requiring translation, scaling, and rotation of the image (Olshausen, Anderson, & Essen, 1995). Deneve acknowledges however that one of the main drawbacks is the curse of dimensionality as the number of basis functions required to approximate functions with high accuracy increases exponentially with the number of signals being combined: N^2, N^3, ..., (Deneve & Pouget, 2003). Deneve further argues in Deneve and Pouget (2003) that this weakness of the basis function approach is also one of its main strengths. Because it uses so many units, a basis function representation tends to be highly redundant. This redundancy can be exploited to filter out the noise optimally.

One question that remains is how our framework can permit to describe an accurate geometrical information about the world. For instance, infants do seem to have an intuition of euclidean geometry that they retrieve from their senses (Spelke, Lee, & Izard, 2010). Adults perform easily mental rotations by simulating how the object would move by applying affine-like transforms (Kosslyn

et al., 2006). One possible hypothesis that we can formulate could come from the findings of "cosine" gain-modulated neurons in the cortical motor areas and responding to certain visual direction and arm movements for performing 3D reaches (Blohm & Crawford, 2012; Blohm et al., 2008; Kakei, Hoffman, & Strick, 2003). We can hypothesize for instance that such neurons may serve for extracting geometrical rules based on our actions (Izard, Pica, Dehaene, Hinchey, & Spelke, 2011). We let nonetheless this unsolved question for future works.

References

Andersen, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *353*, 1421–1428.

Andersen, R., & Buneo, C. (2002). Intentional maps in posterior parietal cortex. *Annual Review of Neuroscience*, *25*, 189–220.

Andersen, R., & Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron*, *63*, 568–583.

Andersen, R., Essick, G., & Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, *230*, 450–458.

Andersen, R., & Mountcastle, V. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *The Journal of Neuroscience*, *3*, 532–548.

Andry, P., Gaussier, P., & Nadel, J. (2002). From visuo-motor development to low-level imitation, In C. G. Prince, Y. Demiris, Y. Marom, H. Kozima, & C. Balkenius, (Eds.) *EpiRob conference* (pp. 1–9).

Andry, P., Gaussier, P., Nadel, J., & Hirsbrunner, B. (2004). Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms. *Adaptive Behavior*, *12*(2), 117–140.

Blohm, G., & Crawford, J. (2009). Fields of gain in the brain. *Neuron*, *64*, 598–600.

Blohm, G., & Crawford, J. (2012). Computations for geometrically accurate visually guided reaching in 3-d space. *Journal of Vision*, *7*(5), 1–22, 4, in special issue 'Sensorimotor processing of goal-directed movements'.

Blohm, G., Khan, A., & Crawford, J. (2008). Spatial transformations for eye-hand coordination. *New Encyclopedia of Neuroscience*, *9*, 203–211.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*, 489–495.

Braun, D., Aertsen, A., Wolpert, D., & Mehring, C. (2009). Motor task variation induces structural learning. *Current Biology*, *19*, 352–357.

Braun, D., Mehring, C., & Wolpert, D. (2010). Structure learning in action. *Behavioural Brain Research*, *206*, 157–165.

Bremner, L., & Andersen, R. (2012). Coding of the reach vector in parietal area 5d. *Neuron*, *75*, 342–351.

Bremner, A., Holmes, N., & Spence, C. (2008). Infants lost in (peripersonal) space? *Trends in Cognitive Sciences*, *12*(8), 298–305.

Chinellato, E., Antonelli, M., Grzyb, B., & del Pobil, A. (2011). Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Transactions on Autonomous Mental Development*, *7*(3), 43–53.

Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, *12*(4), 577–596.

Del Giudice, M., Manera, V., & Keysers, C. (2008). Programmed to learn? The ontogeny of mirror neurons. *Developmental Science*, *104*(5), 1726–1731.

Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron*, *37*, 347–359.

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology—Paris*, *98*, 249–258.

De Rengervé, A., Boucenna, S., Andry, P., & Gaussier, P. (2010). Emergent imitative behavior on a robotic arm based on visuo-motor associative memories. In *International conference on intelligent robots and systems, IROS, 2010, Taipei, Taïwan*.

Fogassi, L., Ferrari, P., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*, 662–667.

Fuke, S., Ogino, M., & Asada, M. (2009). Acquisition of the head-centered peri-personal spatial representation found in VIP neuron. *IEEE Transactions on Autonomous Mental Development*, *1*, 131–140.

Galeazzi, J., Mender, B. M. W., Paredes, M., Tromans, J., Evans, B., Minini, L., & Stringer, S. (2013). A self-organizing model of the visual development of hand-centred representations. *PLoS One*, *8*(6), e66272.

Gallese, V., & Sinigaglia, C. (2011). Neurophysiological mechanisms underlying the understanding and imitation of action. *Trends in Cognitive Sciences*, *15*(11), 512–519.

Goldenberg, G., & Iriki, A. (2007). From sticks to coffee-maker: mastery of tools and technology by human and non-human primates. *Cortex*, *43*, 285–288.

Halgand, C., Soueres, P., Trotter, Y., Celebrini, S., & Jouffrais, C. (2010). A robotics approach for interpreting the gaze-related modulation of the activity of pre-motor neurons during reaching. In *IEEE International conference on biomedical robotics and biomechatronics, BioRob 2010, Tokyo, Japan, 26/09/2010–29/09/2010* (pp. 728–733).

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*, 253–261.

Hoffmann, M., Marques, H., Arieta, A., Sumioka, M., Lungarella, H., & Pfeifer, R. (2010). Body schema in robotics: a review. *IEEE Transactions in Autonomous Mental Development*, *2*(4), 304–324.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330.

Izard, V., Pica, P., Dehaene, S., Hinchey, D., & Spelke, E. (2011). Geometry as a universal mental construction. *Space, Time and Number in the Brain*, *19*, 319–332.

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, *14*(1), S103–S109.

Kakei, S., Hoffman, D., & Strick, P. (2003). Sensorimotor transformations in cortical motor areas. *Neuroscience Research*, *46*, 1–10.

Keysers, C. (2004). Demystifying social cognition: Hebbian perspective. *Trends in Cognitive Sciences*, *8*, 501–507.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*, 635–642.

Kosslyn, S., Ganis, G., & Thompson, W. (2006). Mental imagery and the human brain. In Q. Jn, M. R. Rosenzweig, G. d'Ydewalle, & H. Zhang (Eds.), *Progress in psychological science around the world, neural, cognitive developmental issues. Vol. 1* (pp. 195–209). New York: Psychology.

Kuniyoshi, Y., Yorozu, Y., Inaba, M., & Inoue, H. (2003). From visuo-motor self learning to early imitation—a neural architecture for humanoid learning. *International Conference on Robotics and Automation*, 3132–3139.

Lewkowicz, D., Delevoye-Turrell, Y., Bailly, D., Andry, P., & Gaussier, P. (2013). Reading motor intention through mental imagery. *Adaptive Behavior*, *21*(5), 315–327.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79–86.

McGuire, L., & Sabes, P. (2009). Sensory transformations and the use of multiple reference frames for reach planning. *Nature Neuroscience*, *12*(8), 1056–1061.

Meltzoff, A. (2007). 'Like me': a foundation for social cognition. *Developmental Science*, *10*(1), 126–134.

Meltzoff, A., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuro-science. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *358*, 491–500.

Miyazaki, M., & Hiraki, K. (2006). Delayed intermodel contingency affects young children's recognition of their current self. *Child Development*, *77*, 736–750.

Nagai, Y., Kawai, Y., & Asada, M. (2011). Emergence of mirror neuron system: Immature vision leads to self-other correspondence. In *Proceedings of the first joint IEEE international conference on development and learning and on epigenetic robotics* (pp. 1–6).

Nagy, E., & Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior and Development*, *27*, 54–63.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464.

Okanoya, K., Tokimoto, N., Kumazawa, N., Hihara, S., & Iriki, A. (2008). Tool-use training in a species of rodent: the emergence of an optimal motor strategy and functional understanding. *PLoS One*, *3*(3), e1860.

Olshausen, C., Anderson, B., & Essen, D. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representations. *Journal of Computational Neuroscience*, *2*, 45–62.

Pearson, A., Ropar, D., & Hamilton, A. de C. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience*, *7*(652), 1–10.

Pitti, A., Blanchard, A., Cardinaux, M., & Gaussier, P. (2012). Gain-field modulation mechanism in multimodal networks for spatial perception. In *12th IEEE-RAS International conference on humanoid robots November 29–December 1, 2012. Business Innovation Center Osaka, Japan* (pp. 297–302).

Pitti, A., Braud, R., Mahé, S., Quoy, M., & Gaussier, P. (2013a). Neural model for learning-to-learn of novel task sets in the motor domain, Frontiers in Psychology 4 (771).

Pitti, A., Kuniyoshi, Y., Quoy, M., & Gaussier, P. (2013b). Modeling the minimal new-born's intersubjective mind: the visuotopic-somatotopic alignment hypothesis in the superior colliculus. *PLoS One*, *8*(7), e69474.

Pitti, A., & Kuniyoshi, Y. (2012) Neural models for social development in shared parieto-motor circuits. Book Chapter 11 in "Horizons in Neuroscience Research. Volume 6", Nova Science Publishers, pp.247–282.

Pitti, A., Mori, H., Kouzuma, S., & Kuniyoshi, Y. (2009). Contingency perception and agency measure in visuo-motor spiking neural networks. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 86–97.

Pitti, A., Mori, H., Yamada, Y., & Kuniyoshi, Y. (2010). A model of spatial development from parieto-hippocampal learning of body-place associations. In *10th International conference on epigenetic robotics* (pp. 89–96).

Pouget, A., & Snyder, L. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, *3*, 1192–1198.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–141.

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

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a pet investigation of agency. *Nature Neuroscience*, *4*(7), 546–550.

Salinas, E., & Sejnowski, T. J. (2001). Gain modulation in the central nervous system: Where behavior, neurophysiology and computation meet. *The Neuroscientist*, *7*, 430–440.

Salinas, E., & Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron*, *27*, 15–21.

Schwabe, L., & Blanke, O. (2007). Cognitive neuroscience of ownership and agency. *Consciousness and Cognition*, *16*(3), 661–666.

Shadmehr, R., & Wise, S. (2005). *The computational neurobiology of reaching and pointing: a foundation for motor learning*, MIT Press.

Shimada, S., & Hiraki, K. (2006). Infant's brain responses to live and televised action. *NeuroImage*, *32*(2), 930–939.

Shimada, S., Hiraki, K., & Oda, I. (2005). The parietal role in the sense of self-ownership with temporal discrepancy between visual and proprioceptive feedbacks. *Neuroimage*, *24*, 1225–1232.

Spelke, E., Lee, S., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, 1–22.

Surtees, A., Apperly, I., & Samson, D. (2013). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience*, *7*(698), 1–12.

Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, *14*, 715–725.

Tsakiris, M., Prabhu, G., & Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Consciousness and Cognition*, *15*(2), 423–432.

Van Rullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *BioSystems*, *48*, 229–239.

Van Rullen, R., & Thorpe, S. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*, 2593–2615.