View Sensitive Cells as a Neural Basis for the Representation of Others in a Self-Centered Frame of Reference

Eric L. Sauser*
*Ecole Polytechnique Fédérale de Lausanne
Swiss Federal Institute of Technology, Lausanne
Autonomous Systems Laboratory
CH-1015 Lausanne
eric.sauser@epfl.ch

Aude G. Billard†
†Ecole Polytechnique Fédérale de Lausanne
Swiss Federal Institute of Technology, Lausanne
Autonomous Systems Laboratory
CH-1015 Lausanne
aude.billard@epfl.ch

Abstract

This work follows from a research project, in which we investigate the underlying mechanisms of human imitation and develop a neural model of its core neural circuits. The present paper presents a model of a neural mechanism by which an imitator agent can map movements of the end effector performed by other agents onto its own frame of reference. The model mechanism is validated in simulation and in a humanoid robot to perform a simple task, in which the robot imitates movements performed a human demonstrator.

1 Introduction

Imitation is the ability to recognize, learn and reproduce others' actions. This powerful cognitive mechanism is fundamental for the transmission of knowledge and skills within the same species and across species. It is also at the basis of primates' social communication. One can distinguish the numerous forms of imitation behavior displayed in nature according to levels of complexity. In its simplest form, imitation can be reduced to a sensorimotor mapping that transforms sensory information, usually visual, into corresponding motor commands. Such basic imitation would be displayed as a form of "emulation" or of "social facilitation" (Heyes, 2001). Moreover, Bekkering et al. (2000) have shown that imitation is generally goal-directed, that is children and adults tend more to reproduce the goal of an demonstrative act, rather that the exact sequence of movements leading to it. Indeed, compared to "mimicry", this mechanism doesn't require any body correspondence. *What* to imitate preponderates on *how* to imitate. Then, in its most complex form, imitation leads to or requires more complex cognitive capabilities, such as the recognition of conspecifics and the attribution of others' intentions or states of mind (Billard, 2002). It is often referred to as "true imitation" i.e. the ability to reproduce and learn new motor skills which are not part of the imitator's current motor repertoire. In true imitation, the imitator must be capable to extract the purpose of a given sequence of movements, namely to be capable of action understanding[1].

In this paper, we aim at exploring the mechanisms underlying mimicry. Despite not being directly involved in the most common imitation mechanism that is goal-directed imitation, it is important to note, as mentioned by Wholschläger et al. (2003) that

> *it seems that if the goal is clear (or absent), then the course of the movement plays a more central role in imitation. One might also say therefore, that the movement itself becomes the goal.*

Thus, the question we will develop here is how one can map motions performed by others onto his/her own perspective, and more precisely while considering the end effector trajectory, i.e. the hand of the demonstrator. Indeed, as simple as it appears to be to Ethologists, mimicry remains complex in terms of the basic cognitive capabilities it requires, such as the capacity to perform arbitrary frames of references transformations and to generate coherent sensorimotor mappings. Such cognitive processes are fundamental and necessary for more complex forms of imitation. They remain, however, ill-understood. We argue that a better understanding of the brain mechanisms underlying mimicry is necessary to provide the stages for understanding and modeling the leap from simple to complex forms of imitation in animals.

While the behavioral processes of imitation have been the focus of studies in Ethology and develop-

[1] An action is understood here as a goal-directed sequence of movements.

mental Psychology for centuries, evidence of correlated neural processes is much more recent. Studies of brain lesions resulting in degenerate imitative behaviors (apraxia or echopraxia) were the first to give some insight into the brain areas responsible for imitation, pointing to generic areas in the frontal and parietal cortices (Lhermite et al., 1986; Shimomura and Mori, 1998). The field revived a new life with the discovery in 1992 of the *mirror neurons* system, direct-mapping mechanism between visual and motor systems. For recall, mirror cells respond both when the animal performs and sees a goal-directed sequence of movements, hence, suggesting that a direct-mapping mechanism between visual and motor system exists for the purpose of linking conspecifics' or humans' action observation with self motor execution. The mirror neurons were first detected in the macaque monkey premotor cortex (PM), posterior parietal cortex (PPC) and superior temporal sulcus (STS) (Fogassi and Gallese, 2002; Rizzolatti et al., 1996). Later, brain imaging studies of the human brain highlighted numerous areas, such as STS, PM and Broca (Nishitani and Hari, 2000; Iacoboni et al., 1999; Decety et al., 2002). While the discovery of the mirror neuron system is certainly a key step toward a better understanding of the brain mechanisms underlying primates' ability to perform various forms of imitation, one has yet to clearly spell out the role of the mirror neuron system as part of the general neural processes for imitation.

Mirror neurons are relatively far from the brain areas receiving primary sensory information. They react, thus, to highly processed stimuli, represented in a goal-centered frame of reference (FR). A proposal by Burnod et al. (1999) suggested that the series of FRs, required for transferring information in retina-based FR into a body-centered FR, is encoded by different cells along the visual pathway, following a sensory gradient of increasing complexity. Indeed, along the visual pathway (the "what stream"), the information flows from the primary visual cortex (V1) to the temporal lobes, including the inferior temporal area (IT) and the superior temporal sulcus(STS). IT contains populations of neurons that separately exhibit sensitivity to a variety of objects. Some of these populations are sensitive to the size and orientation relative to an viewer-centered FR, whereas others react in an object-centered FR (Booth and Rolls, 1998). Similarly, neurons in macaques' STS, have been found to respond to specific human body parts and correlate with various quantities such as the position, rotation and translation of limbs, hands, faces, eyes; as well as with complex motions such as walking. Perrett et al. (1989) showed that the FRs in which these neurons seem to react are multiple. Moreover, there is a body of evidence that spatial visual properties such as direction, orientation and size of objects are also encoded in PPC (Sakata et al., 1999). Finally, concerning the distance of the target objects and observed bodies, neurons activities in the ventral pathway and parietal cortex have been shown to correlate this parameter, firing differently for close or far stimuli in a modulatory fashion (Dobbins et al., 1998; Sakata et al., 1980). All these regions are tightly coupled and form a complex network (Wise et al., 1997) that plays a fundamental role in primates ability to reproduce movements and goal-directed actions, such as transforming viewer-centered information into an other-centered representation.

In former work (Arbib et al., 2000; Billard and Matarić, 2001), we started developing computational models of the complete visuomotor pathway underlying imitative behavior. In this paper, we present a neural model that accounts for the ability to perform arbitrary frame of reference transformations and to display mimicry of hand motions. The model attempts, once the goal has been clearly identified, to explain the core circuits underlying the ability to map goal-directed motion performed by others into a frame of reference located onto one's own body. Such basic imitative behavior is displayed both by monkeys and humans.

2 A Mimicry Task

The mimicry task we consider in this paper is illustrated in Figure 1. It consists of the following: An imitator and a demonstrator face one another. The demonstrator produces various movements with his right hand. The imitator tries to reproduce the demonstrator's actions simultaneously (immediate imitation). The imitator attempts to reach to the same location as the demonstrator's hand in its own frame of reference. For instance, when the demonstrator's hand performs a circular trajectory on his left side, the imitator has to perform a similar hand motion on his **own left**, independently on the demonstrator's orientation. Indeed, the imitator could face the observer, be on a profile view or even be turned upside-down. It must be able to still perform the correct frame of reference transformation.

However simple this task appears to be, it is non trivial to model the neural processes that underly it. Thus, we describe a distributed neural model, inspired from neurophysiological evidence of population vector coding, that is able to perform such

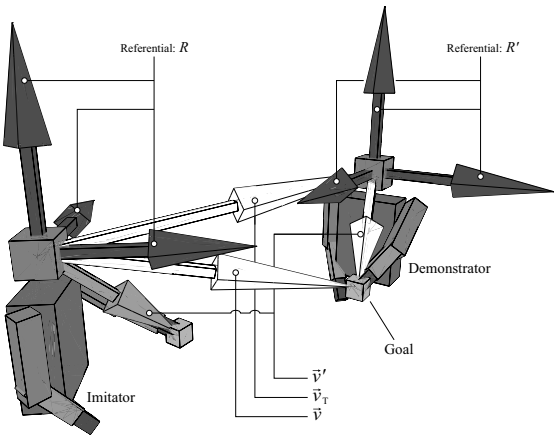

Figure 1: Illustration of the frames of reference transformation required to transfer the target from the demonstrator's view point to that of the imitator.

transformations. As an illustration, we implement this mechanism into a robotic platform using a mini-humanoid robot shown in Figure 8, to perform immediate imitation of hand drawings produced by a human demonstrator.

2.1 The Frames of Reference Problem

Consider the core problem tackled in this paper: How does the central nervous system perform frames of references transformation in order to build a body-centered or object-centered representation?

In mathematical terms, as illustrated on Figure 1, the question is how can we transform a vector $\vec{v}$ given in a referential R into $\vec{v}\,'$ in R', knowing the vector $\vec{v}_{\mathrm{T}}$ across the origins of the two referentials, and the axes of the referential R' itself, expressed in R. We assume that R and R' are given by

$$\begin{aligned} R &= \{O, \vec{e}_1, \vec{e}_2, \vec{e}_3\} \\ R' &= \{O', \vec{e}\,'_1, \vec{e}\,'_2, \vec{e}\,'_3\} \end{aligned}$$

where $OO' = \vec{v}_{\mathrm{T}}$, and $\vec{e}\,'_i$, $\vec{e}_i$, $\forall i \in \{1..3\}$ correspond to the principal axes, as unit vectors, of the demonstrator's body and of the observator's body, respectively. These axes correspond to the right-left, feet-head and back-front axes, respectively. The orientation of R' with respect to R is given by the rotation matrix $M_{R'}$:

$$M_{R'} = \left(\begin{array}{c|c|c} \vec{e}\,'_1 & \vec{e}\,'_2 & \vec{e}\,'_3 \end{array} \right). \qquad (1)$$

By writing down the classical transformations across referentials and considering M_R as an identity matrix, we get the following forward and inverse equations:

$$\begin{aligned} \vec{v} &= M_{R'}\vec{v}\,' + \vec{v}_{\mathrm{T}} \\ &\Leftrightarrow \\ \vec{v}\,' &= M_{R'}^{-1}\left(\vec{v} - \vec{v}_{\mathrm{T}}\right). \end{aligned} \qquad (2)$$

If we consider now that $M_{R'}$ is orthonormal, we know that $M_{R'}^{-1} = M_{R'}^{T}$. This allows us to rewrite the previous equation using the dot product and we find:

$$\vec{v}\,' = \sum_{i \in \{1..3\}} \left(\vec{e}\,'_i \cdot \left(\vec{v} - \vec{v}_{\mathrm{T}}\right)\right) \vec{e}_i. \qquad (3)$$

Such transformation can thus be reduced to a combination of relatively simple (from a neurophysiological point of view) vectorial operations, consisting of sums, dot products, and unitary vector scaling.

This way, the vector $\vec{v}\,'$ pointing to the target in the demonstrator's referential can be directly mapped into the imitator-centered referential, so that the demonstrator's target is considered as the imitator's one.

2.1.1 Population Vector Coding

We use the *population vector coding* paradigm as a neurophysiological substrate for representing each of the vectors of our referentials.

In this paper, we define a population as an ensemble of neurons whose distributed firing activities are correlated to a single macroscopic quantity that is a vector $\vec{v}$ in a given frame of reference R. In such populations, each neuron is tuned to a preferred direction $\vec{r}$, i.e. its firing activity is maximal when $\vec{v}$ and $\vec{r}$ are collinear and point to the same direction, and decrease as $\vec{v}$ diverges from $\vec{r}$. Then, in order to extract the information from a populations of neurons, as originally proposed by Georgopoulos (1996), we use the population vector. Considering that each neuron votes for its preferred direction proportionally to its firing activity, by taking the average of all these votes, we obtain the vector encoded by this population, i.e. the population vector.

2.1.2 View Sensitive Cells Defining Referentials

As mentioned in Section 1, we know from neurophysiology that neurons in STS and IT are sensitive to different orientations or views of bodies and objects, respectively. These neurons firing activities have also been shown to be correlated to different frames of references, mainly in a viewer, object or goal centered reference frame. Moreover, these populations

of neurons exhibit a large range of preferred directions, tending to cover uniformly the possible orientation given rotations around the three principal axis.

In order to model these cells, we assume that there are three distinct populations of neurons encoding separately the three principal axis of a observed body or object. This principle might be consistent with neurophysiological data despite not being completely experimentally proved. Indeed, to our knowledge, there are no neurophysiological experiments that have systematically tested the response of orientation sensitive cells to the complete ensemble of possible orientation. Indeed, usually sole the classical rotations along the three principal axis were tested.

2.2 The Model

This section presents a summary of a neural model for frames of reference transformations that we have, in its major parts, already proposed in Sauser and Billard (2005). For more information on the mathematical development and on the implementation details, please refer to this paper. The novelty here, concerns the parallel use of three principal axes determining a frame of reference, rather than a set of angles that code for a series of rotations that are performed serially.

2.2.1 An attractor network

Let us consider Ω, a continuous population of neurons where each unit participating in the population is characterized by its preferred direction $\vec{r}$. In this paper, the preferred directions are assumed to be uniformly distributed along a 3 dimensional subspace $\Gamma = \{\vec{r} \in \mathbb{R}^3 \mid \|\vec{r}\| = 1\}$, that corresponds to the surface of a unitary sphere. The response of the whole population, the population vector, is given in a continuous form by

$$\vec{P} = \frac{1}{\kappa} \oint_{\Gamma} f\left(u_{\vec{r}}\right) \vec{r} \, \mathrm{d}\vec{r} \tag{4}$$

where $\kappa = \frac{2\pi}{3}$ is a normalization factor, $u_{\vec{r}}$ the neuron's membrane potential with preferred direction $\vec{r}$, and $f(u_{\vec{r}})$ its firing activity. f is a non-linear function equal to $f(x) = \max(0, x)$.

Let us now consider an attractor network (Salinas and Abbott, 1996) made of a fully connected population of neurons whose dynamics is governed by

$$\begin{aligned} \tau \dot{u}_{\vec{r}} &= -u_{\vec{r}} + \oint_{\Gamma} w_{\vec{r}' \to \vec{r}} \, f(u_{\vec{r}'}) \, \mathrm{d}\vec{r}' + x_{\vec{r}} \\ w_{\vec{r}' \to \vec{r}} &= \gamma(\eta) \left(\vec{r}' \cdot \vec{r}\right) \end{aligned} \tag{5}$$

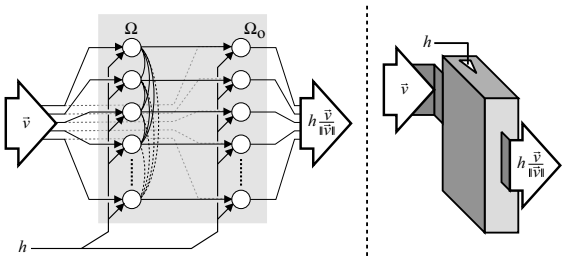

Figure 2: On the left, architecture of the two layers neural network producing a non-linear composition of its inputs. On the right, the symbolic illustration of this network as will be used further in the paper.

where $w_{\vec{r}' \to \vec{r}}$ are the lateral weights that exhibit symmetric, rotation invariant, and center surround excitation inhibition characteristics, $x_{\vec{r}}$ is the external synaptic input, and $\gamma(\eta)$ is a scaling factor depending on the network parameter $\eta \in]0, 1[$ that controls the influence of the lateral weights[2]. Assuming that the network input $x_{\vec{r}}$ is composed of a vectorial and a constant homogeneous input of the form given by

$$\begin{aligned} x_{\vec{r}} &= \vec{r} \cdot \vec{v} + h \\ &= \beta_v \left(\vec{r} \cdot \vec{r}_v\right) + h \end{aligned} \tag{6}$$

where $\beta_v = \|\vec{v}\|$ and $\vec{r}_v = \frac{\vec{v}}{\|\vec{v}\|}$. We have shown that the activity profile of this network converges toward a stable state that can be approximated by

$$u^{\star}_{\vec{r}} \approx h + \frac{1}{\chi(\eta)} \left(\vec{r} \cdot \vec{v}\right) + \frac{1}{\eta} h \left(\vec{r} \cdot \vec{r}_v\right) \tag{7}$$

where $\chi(\eta) = 1 - \gamma(\eta)\frac{\pi}{3}$. We can see that the approximation of this activity profile reflects both the vectorial and constant inputs, plus a modulatory term, which is the result of the interactions of the recurrent connectivity.

2.2.2 A Two Layers Neural Network

As seen on Equation 7, the current attractor network produces, as an output, a sum of vectorial and constant terms. In order to strictly keep the multiplicative term and thus have a network capable of producing a non-linear composition of its two input sources, we build a two layers neural architecture as illustrated in Figure 2. The first layer consists of the attractor network. The second layer is composed of another population Ω_{O}, o for output, without lateral weights. It receives projections from the recurrent population

[2] In the present case of populations representing 3D vectors, $\gamma(\eta) = \left(\frac{\pi}{3}\left(2 + 3\,\eta - \eta^3\right)\right)^{-1}$, such that Equation (5) has a non trivial state of convergence.

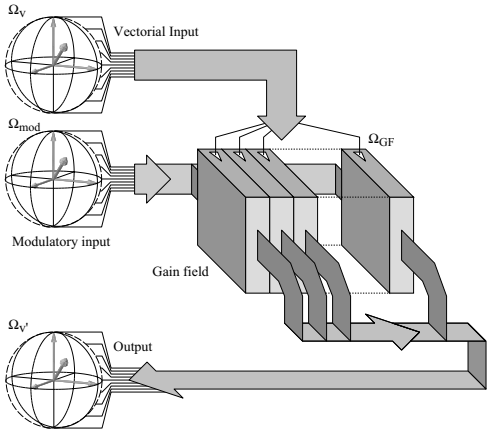

Figure 3: The architecture and connectivities of the gain field. The spheres containing a referential and a vector correspond to populations of neurons coding for a vector in a given referential.

using one to one synapses, and inhibitory inputs corresponding to the vectorial and constant inputs of the recurrent population with an appropriate scaling, such that

$$x^{\mathrm{O}}_{\vec{r}} = \eta\big(f(u_{\vec{r}}) - h - \frac{1}{\chi(\eta)}(\vec{r}\cdot\vec{v})\big). \qquad (8)$$

Considering that the neurons of Ω_{O} support an immediate integration mechanism such that $u^{\mathrm{O}}_{\vec{r}} = x^{\mathrm{O}}_{\vec{r}}$, we obtain, after a substitution in the previous equation using (7), that the firing rate converges toward

$$\begin{aligned} f(u^{\mathrm{O}}_{\vec{r}}) &= f\left(\eta\big(f(u_{\vec{r}}) - h - \frac{1}{\chi(\eta)}\beta_v(\vec{r}\cdot\vec{r}_v)\big)\right) \\ &\approx \begin{cases} h\,(\vec{r}\cdot\vec{r}_v) & \vec{r}\cdot\vec{r}_v > 0,\ h > 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \qquad (9)$$

This network is capable of encoding independently two separate quantities, that are the direction $\vec{r}_v$ and the amplitude h, regardless of the intensity of the directional input β_v. In vectorial terms, this means that given a vector $\vec{v}$ and a scalar h, the output population vector will tend toward $h\frac{\vec{v}}{\|\vec{v}\|}$. Therefore, this model can be used to form a vectorial basis, inspired from classical linear algebra. Moreover, it will also be the building block of a bigger network, the gain field.

2.2.3 The Gain Field

In order to combine two different sources of vectorial information, we propose a model of gain field that follows an architecture and connectivity shown in Figure 3. It consists in an assembly of building blocks, described in Section 2.2.2, that define a new dimension denoted by $\vec{s} \in \Gamma$. The external inputs come from two different vectorial sources represented by a modulatory population Ω_{mod} and an vectorial population Ω_{v}, that encode the vectors $\vec{v}_{\mathrm{mod}}$ and $\vec{v}$, respectively. They are separately applied to each dimension of the gain field, $\vec{r}$ and $\vec{s}$, respectively. Hence, the input for each neuron $\vec{r}$ of each layer $\vec{s}$ in the gain field Ω_{GF} is defined by

$$\begin{aligned} x^{\mathrm{GF}}_{(\vec{r},\vec{s})} &= \oint w^{\mathrm{mod}\to\mathrm{GF}}_{\vec{r}'\to\vec{r}}\, f(u^{\mathrm{mod}}_{\vec{r}'})\,\mathrm{d}\vec{r}' + \\ &\quad \oint w^{\mathrm{v}\to\mathrm{GF}}_{\vec{r}'\to\vec{s}}\, f(u^{\mathrm{v}}_{\vec{r}'})\,\mathrm{d}\vec{r}' \\ &= (\vec{r}\cdot\vec{v}_{\mathrm{mod}}) + (\vec{s}\cdot\vec{v}) \end{aligned} \qquad (10)$$

Then, if we substitute this equation into (9), we obtain that the gain field output firing activity converges toward

$$\begin{aligned} f(u^{\mathrm{GFO}}_{(\vec{r},\vec{s})}) &\approx \beta_v(\vec{s}\cdot\vec{r}_v)(\vec{r}\cdot\vec{r}_{v_{\mathrm{mod}}}) \\ &\approx (\vec{s}\cdot\vec{v})(\vec{r}\cdot\vec{r}_{v_{\mathrm{mod}}}) \end{aligned} \qquad (11)$$

From this, we can see that the activity profile of the gain field output is symmetric and that the peak is located at the intersection of the directions currently encoded by both source populations. Moreover, considering the amplitude of its activity, sole the amplitude of $\vec{v}$ is taken into consideration in this network. This property allows transformations that guarantee that the amplitude of the transformed vectorial quantity is preserved (Sauser and Billard, 2005).

2.2.4 Projections on Principal Axis and Others Centered Frame of Reference

The final step, and the new part of our model, is to show how our neural network model can perform arbitrary frames of reference transformations by applying the principles mentioned in Section 2.1 (see Equ. (3)). As shown in Figure 4, we consider five sources of information arising from five populations of neurons that encode $\vec{e}\,'_i$, $i \in \{1..3\}$, $\vec{v}$ and $\vec{v}_{\mathrm{T}}$. In order to compute the dot product, we need three gain fields whose modulatory inputs are connected to the populations coding for the principal axis $\vec{e}\,'_{i\in\{1..3\}}$, while their vectorial inputs are linked to the difference between populations coding for $\vec{v}$ and $\vec{v}_{\mathrm{T}}$ that are connected using excitatory and inhibitory synapses, respectively. These gain fields project then to another population that will receive the result of the transformation: the vector $\vec{v}\,'$ in a body or object centered frame of reference using the following synaptic weights, $\forall i \in \{1..3\}$

$$w^{\mathrm{GFO}_i\to\mathrm{v}'}_{(\vec{r},\vec{s})\to\vec{r}\,'} = \frac{1}{\kappa^2}(\vec{r}\cdot\vec{s})\ (\vec{r}\,'\cdot\vec{e}_i). \qquad (12)$$

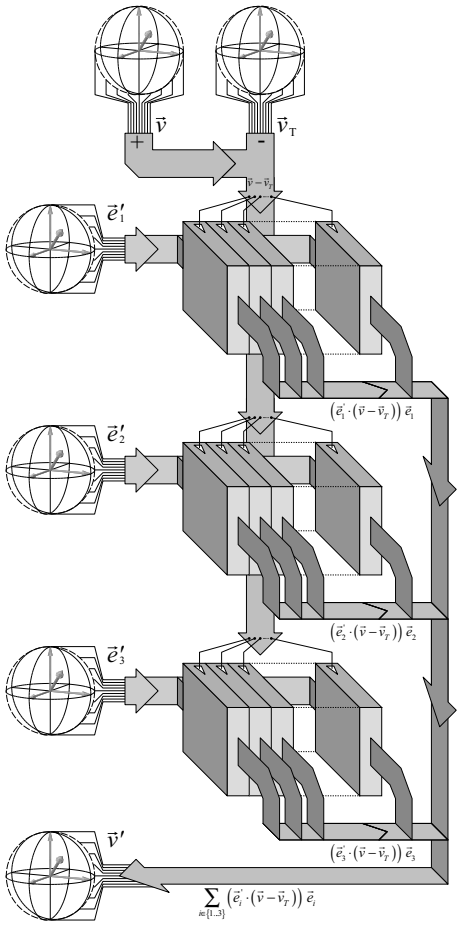

Figure 4: The architecture and connectivities of the model that can perform frames of reference transformations.

Then, using the activity profile of the gain fields described by Equation (11), each neuron of the final population receives a synaptic input equal to

$$\begin{aligned} x^{\mathrm{v}'}_{\vec{r}\,'} &= \sum_{i\in\{1..3\}} \oint f\left(u^{\mathrm{GFO}_i}_{(\vec{r},\vec{s})}\right)\, w^{\mathrm{GFO}_i\to\mathrm{v}'}_{(\vec{r},\vec{s})\to\vec{r}\,'}\, \mathrm{d}\vec{r}\mathrm{d}\vec{s} \\ &= \sum_{i\in\{1..3\}} \left((\vec{v}-\vec{v}_{\mathrm{T}})\cdot\vec{e}\,'_i\right)(\vec{r}\,'\cdot\vec{e}_i) \\ &= \vec{r}\,'\cdot\vec{v}\,' \end{aligned} \tag{13}$$

This equation means that this population is now encoding $\vec{v}\,'$ in a body or object centered frame of reference.

2.3 Experimental Setup

We implemented this system in a kinematic simulator of a pair of demonstrator - imitator humanoid avatars

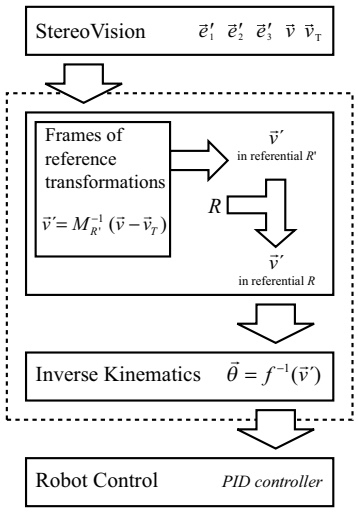

Figure 5: Overview of the system implementation on a robotic platform. The surrounding dotted rectangle indicates the parts used the simulation.

(see Fig. 6) and in a humanoid robot, as shown in Figure 8. An overview of the overall system architecture is illustrated in Figure 5. The visual system consists in two webcams connected to a color-based stereo vision software that allow the tracking of specific colors marks in 3D space. The human demonstrator is placed in front of the cameras, with three different color marks on the left and right of his torso, and on his hand. Assuming that he is always in a standing posture the two marks on the body are sufficient to uniquely determine the demonstrator's principal axis, that are $\vec{e}\,'_i$, $i \in \{1..3\}$. The visual system also provides the body and hand position in a viewer centered frame of reference, $\vec{v}_{\mathrm{T}}$ and $\vec{v}$, respectively. These information are fed into our neural network in order to compute the target location in the demonstrator's body centered reference frame. It is directly applied to a self-centered frame of reference that gives the imitator its own target. In order to allow the robot to reach the target, this position of the target with respect to the imitator is fed to an inverse kinematic algorithm adapted from Wang and Verriest (1998), that provides the sequence of joint angles to the robot.

3 Results

3.1 Mimicry of hand gestures

We conducted simulations, in which the demonstrator avatar draw 8 different figures. Figure 7 shows superimposed the trajectories performed by the demonstrator and the imitator. Demonstrated and imitated

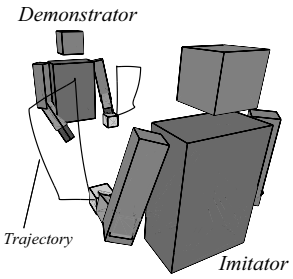

Figure 6: One avatar is drawing a figure while the other imitates the demonstrated trajectory.

movements show a high qualitative resemblance. However, one can observe a systematic shift in space and a slight deformation of the figure. This is an artifact resulting from the non-uniformity of the distribution of preferred directions in our neural population[3]. In other words, the neural populations produce a non-uniform map of their inputs, resulting in a slight deformation of the three dimensional representation of the target vectors.

We, then, conducted experiments, in which a humanoid robot imitated 4 trajectories produced by a human demonstrator. Figure 8 shows, superimposed, the trajectories of the demonstrator's and imitator's hand for the four examples. We can observe that the results are similar to those obtained in simulation. The imitation is qualitatively good. However, it suffers from a systematic shift in space and rescaling in amplitude. In addition, the use of a stereovision system for recording demonstrated and imitated trajectories creates a new source of errors.

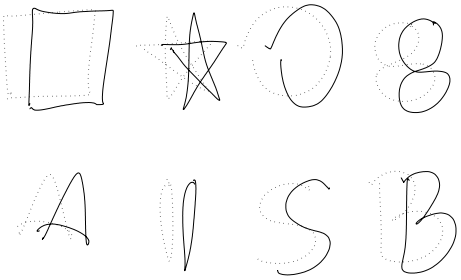

Figure 7: Eight trajectories followed by the demonstrator's hand (dotted lines) and by the imitator's hand (plain lines) in simulation.

[3]In Sauser and Billard (2005), we showed that only a "quasi" uniform distribution of preferred directions can be obtained using iterative algorithms.

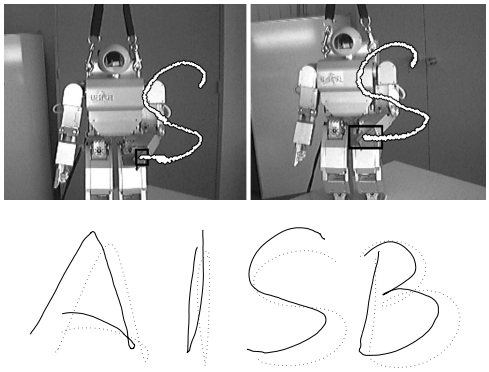

Figure 8: Top Figure: Hoap-2 a mini-humanoid robot built by Fujitsu, provided with 25 degrees of freedom, including 4 on each arm. The robot imitates a human trajectory forming an "S", while tracking the demonstrator's gesture using a pair of fixed cameras. Bottom Figure: Four trajectories followed by the demonstrator's hand (dotted lines) and by the robot's hand (plain lines) in simulation.

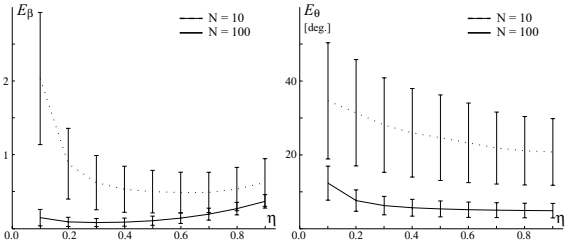

Figure 9: Error recorded during a simulation batch where populations with different parameters were given random input vectors and referentials.

3.2 Error measures

In addition to the errors that appear by discretizing continuous equations, the approximation we made in our mathematical development (see Equ. (7)) is a source of systematic errors between the theoretical resulting vector, denoted by $\vec{v}'^{\star}$, computed with classical algebraic equations, and the result $\vec{v}'$ produced by our network. To quantify them, we define E_β, the error on the amplitude, and E_θ, the error on the direction, by

$$E_\beta(\vec{v}', \vec{v}'^{\star}) = \frac{|\,\|\vec{v}'\| - \|\vec{v}'^{\star}\|\,|}{\|\vec{v}'^{\star}\|} \quad (14)$$

$$E_\theta(\vec{v}', \vec{v}'^{\star}) = \text{acos}\left(\frac{\vec{v}' \cdot \vec{v}'^{\star}}{\|\vec{v}'\|\,\|\vec{v}'^{\star}\|}\right) \quad (15)$$

that correspond to the relative difference between their norms, and to the angle they form, respectively. Figure 9 shows the errors E_β and E_θ that were

measured during a simulation batch where random vectors where transformed into random referentials. The different curves correspond to different network sizes. First, we can see as expected that the bigger is a population, the smaller are the errors. Second, consistently with our previous work (Sauser and Billard, 2005), the parameter η has a ambivalent influence on the network. On the one hand, small values increase the importance of the recurrent connections, hence increasing the errors due to an imperfect distribution of preferred directions. On the other hand, big values induce more errors due by our mathematical approximations (see Equ. (7)). These two properties explain why, on the left of the figure, an optimum can be observed.

4 Discussion and Conclusion

The model, presented in this paper, provides an example of neural mechanism for the representation of others in a self-centered frame of reference. As such, it is an important step toward a full-scale imitation model. Indeed, as illustrated in this paper, a model for solving the frames of reference transformation problem provides us automatically with a simple imitation mechanism. Note that the present model does not yet explain the tendency humans have to perform imitation in mirror fashion when reproducing meaningless gesture, and when demonstrator and imitator face each other (Wholschläger et al., 2003). It only shows a solution to the frame of reference problem. Note that the model could be extended to address this issue. The preference for mirror imitation could simply be an effect of early visual processing, occurring prior to the frame of reference transformation, that would represent the demonstrator's body in a referential that reflect best the natural symmetries of the human body; presenting motions perceived visually on the left handside of the imitator by corresponding motion on the left handside of the imitator.

Another important aspect not yet addressed by our model is how the rescaling of the demonstrator's motions to the imitator's body is performed. In the present implementation, rescaling is done by hand, providing a vector of an appropriate size to the network, so that the resulting vector after convergence lies within the robot's range of motion. The model could be extended to encapsulate explicitly the rescaling aspect, by exploiting the multiplicative nature of the network. Moreover, such a neural representation would be in accordance with biological evidence that neurons located in the visual cortex fire in response to the size of an object, regardless of the distance to the object (Dobbins et al., 1998; Sakata et al., 1980).

There is as yet no evidence to support our model's hypothesis that orientation sensitive cells in the visual areas STS and IT are grouped in populations that encode the principal axes of the demonstrator's body. If evidence of such an encoding was to be found, this would suggest that such groups of neurons may form a basis (in the vectorial sense) of a goal centered representation of hand motion. Unfortunately, to our knowledge, no systematic experiment have shown a complete description of single cell sensitivity to all possible orientations. Note that if these cells were to encode the three principal axes, this would offer a highly redundant representation of motion. One could consider less redundant forms of encoding 3D frames of reference. However, as discussed by Marr (1982), such representations are difficult to determine and the three axes representation remain the most natural representation for 3D frames of reference. Furthermore, Deneve and Pouget (2003) proposed a model that deals with a two dimensional object-centered representation using basis functions. The authors argue that a redundant neural substrate is well-suited to reduce neural noise and to simplify the complexity of single cell computation.

The time required for the model to perform a FR transformation is independent on the orientation of the two frames of reference. Such a result is in contradiction with the observation that humans produce a longer reaction time, when required to perform mental rotations in an "unusual" orientation, such as shifting an image upside-down. One could, however, imagine that another mechanism is at play. In absence of visual input, such a mechanism would set the principal axes of the demonstrator's referential to a default state (i.e. setting the preferred directions of the network in our model to a default value), expressing the imitator's expectation that the demonstration would stand vertically and would face him. In this case, the network's state in our model will take more time to match unusual visual orientations; hence, reproducing the expected observation. Note that the delay could also be due to a longer processing phase during preprocessing of the visual field, for recognizing the body features (used then to set the landmarks for determining the axes).

Finally, we showed that frame of reference transformation performed by the model result in qualitative discrepancies between demonstrated and imitated trajectories, while ensuring a high qualitative resemblance across demonstrated and imitated motions. Note that humans show also imprecision in their imitation, if other constraints, such as an align-

ment to landmarks, are not specified or absent. In future work, we will compare the imprecisions made by the model to those done by human imitators.

The model's implementation we presented in this paper focused on a body-centered frames of reference transformation. The model is, however, quite general and could, also, be applied to object-centered representations. The later representation being crucial to performing several daily tasks. In future work, this model will be adapted to form both object- and goal-centered representations in order to provide context dependent information for goal-directed imitation.

Acknowledgments This work was supported by the Swiss National Science Foundation, through grant no 620-066127 of the SNF Professorships program.

References

M. Arbib, A. Billard, M. Iacoboni, and E. Oztop. Mirror neurons, imitation and (synthetic) brain imaging. In *Neural Networks*, volume 13 (8/9), pages 953–973, 2000.

H. Bekkering, A. Wohlschläger, and M. Gattis. Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology*, 53(1):153–164, 2000.

A. Billard. Imitation: A review. In *The Handbook of Brain Theory and Neural Network. 2nd Edition. Michael A. Arbib (editor)*, pages 566–569. MIT Press, 2002.

A. Billard and M. Matarić. Learning human arm movements by imitation: Evaluation of a biologically-inspired connectionist architecture. *Robotics & Autonomous Systems*, 941:1–16, 2001.

M.C.A Booth and E.T. Rolls. View-invariant representations of familiar objects by neurons in the inferior temporal cortex. *Cerebral Cortex*, 8:510–523, 1998.

Yves Burnod, Pierre Baraduc, Alexandra Battaglia-Mayer, Emmanuel Guigon, Etienne Koechlin, Stefano Ferraina, Francesco Laquaniti, and Roberto Caminiti. Parieta-frontal coding of reaching: an integrated framework. *Experimental Brain Research*, 129:325–346, 1999.

J Decety, T. Chaminade, J Grezes, and A. N. Meltzoff. A PET Exploration of the Neural Mechanisms Involved in Reciprocal Imitation. *Neuroimage*, 15:265–272, 2002.

S. Deneve and A. Pouget. Basis function for object-centered representations. *neuron*, 37:347–359, 2003.

A.C. Dobbins, R.M. Jeo, J. Fiser, and J.M. Allman. Distance modulation of neural activity in the visual cortex. *Science*, 281: 552–555, 1998.

L. Fogassi and V. Gallese. The neural correlates of action understanding in non-human primates. *In Mirror Neurons and the Evolution of Brain and Language, Advances in Consciousness Research, John Benjamins Publishing*, pages 13–35, 2002.

A.P. Georgopoulos. On the translation of directional motor cortical commands to activation of muscles via spinal interneuronal systems. *Cognitive Brain Research*, 3:151–155, 1996.

C. Heyes. Causes and conscequence of imitation. *Trends in Cognitive Sciences*, pages 253–261, 2001.

M. Iacoboni, R.P. Woods, M. Brass, H. Bekkering, J.C. Mazziotta, and G Rizzolatti. Cortical mechanisms of human imitation. *Science*, 286:2526–2528, 1999.

F. Lhermite, B. Pillon, and M. Serdaru. Human autonomy and the frontal lobes, par i: imitation and utilization behavior: a neuropsychological study of 75 patients. *Annual Neurology*, 19:326–34, 1986.

D. Marr. Vision. *Cambridge MA: MIT Press*, 1982.

N. Nishitani and R. Hari. Temporal dynamics of cortical representation for action. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 97(2), pages 913–918, 2000.

D. I. Perrett, M Harries, A.J. Mistlin, and A.J. Chitty. Three stages in the classifi cation of body movements by visual neurons. In H.B et al. Barlow, editor, *Images and Understanding*, pages 94–107. Cambridge University Press, 1989.

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

H. Sakata, H. Shibutani, and K. Kawano. Spatial properties of visual fi xation neurons in posterior parietal association cortex of the monkey. *Journal of Neurophysiology*, 43:1654–1672, 1980.

H. Sakata, M. Taira, M. Kusunoki, A. Murata, K. Tsutsui, Y. Tanaka, W.N. Shein, and Y. Miyashita. Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research*, 128:160–169, 1999.

E. Salinas and L.F. Abbott. A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences USA*, 93:11956–11961, 1996.

E. Sauser and A. Billard. Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomputing, In press*, 2005.

T. Shimomura and E. Mori. Obstinate imitation behaviour in differentiation of frontotemporal dementia from alzheimer's disease. *Lancet*, 352(9128):623–4, 1998.

X.G. Wang and J.P. Verriest. A geometric algorithm to predict the arm reach posture for computer-aided ergonomic evaluation. *Journal of Visualization and Computer Amination*, 9:33–47, 1998.

A. Wholschläger, M. Gattis, and H. Bekkering. Action generation and action perception in imitation: an instance of the ideomotor principle. *Phil. Trans. R. Soc. Lond.*, 358:501–515, 2003.

S.P. Wise, D. Broussaoud, P.B. Johnson, and R. Caminiti. Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annual Reviews in Neuroscience*, 20:25–42, 1997.