Abstract for the Computational Neuroscience Meeting, CNS*06, Madison, USA

Neural Model of the Transformation from Allo-centric to Ego-centric Representation of Motions: From Observing Others to Controlling Oneself

Eric L. Sauser and Aude G. Billard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Swiss Federal Institute of Technology, Lausanne
Autonomous Systems Laboratory
CH-1015 Lausanne, Switzerland

In this work, we aim at exploring the mechanisms underlying simple forms of imitation such as mimicry. Specifically, we focus on the problem of how to map an allocentric representation of motions performed by others onto an egocentric representation of self-generated motions, as illustrated in Figure 1.

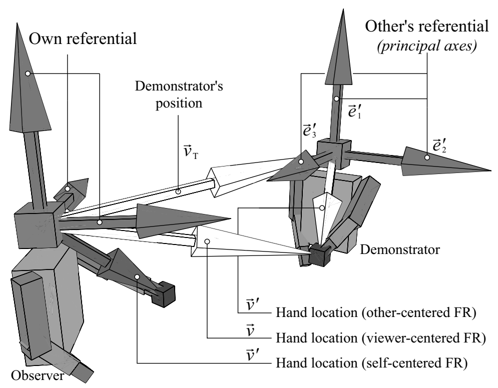


Fig. 1. Illustration of the frames of reference transformations problem.

While considering the neural correlates related to imitative behaviors, the discovery in the monkey brain of the mirror neurons system (MNS) has suggested the existence of a direct-mapping mechanism between visual and motor systems [4]. This link between action observation with self motor execution underlies a strong need for a common visuomotor representation, a common frame of reference (FR). Thus, this naturally raised the question of what might be the neural processes that allow the brain to express visually perceived human motions into an egocentric frame of reference. From neurophysiology, we know, that along the ventral visual pathway, the information flows from the primary visual cortex to the superior temporal sulcus (STS). This region contains populations of neurons that separately exhibit sensitivity to a variety of body parts, and also to their locations, sizes and orientations relative to a viewer, object or goal-centered FR [3]. Therefore, being indirectly connected to the MNS, STS appears clearly to be a candidate where this viewer to body-centered transformation occurs. Our model, that will be described in the next paragraphs, proposes a biologically plausible mechanism of how such transformation might be performed in this brain region.

The Model

The model exploits the *population vector coding paradigm* to represent the vectorial basis of the referentials involved in the transformation. We consider 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. In order to build a neural model of a body-centered frame of reference, we propose the hypothesis that orientation sensitive cells in the visual area STS are grouped into populations that encode the principal axes of observed bodies, as it remains the most natural representation for three dimensional frames

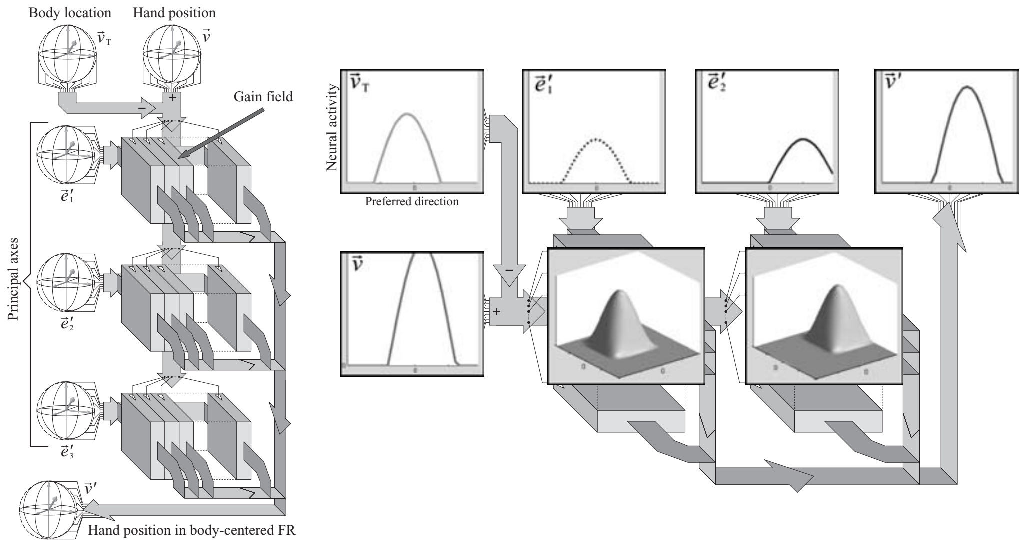


Fig. 2. On the left, architecture of the model that performs arbitrary 3D frames of reference transformations, given three principal axis. On the right, the dynamics of the network are illustrated in a two dimensional case (for clarity reasons).

of reference [2]. Therefore, we consider three distinct populations of neurons for coding separately the three principal axes of an observed body. As such, these groups of neurons can form a basis, in the vectorial sense, of a body-centered frame of reference. Despite there is as yet no clear evidence to support our model's hypothesis, this principle is consistent with current neurophysiological data. Indeed, to our knowledge, no systematic experiment have shown a complete description of single cell sensitivity to all possible orientations.

Formally, we consider a continuous population of neurons where each unit is characterized by its preferred direction $\vec{r}$ that are assumed to be uniformly distributed along a three dimensional subspace $\Gamma = \{\vec{r} \in \mathbb{R}^3 \mid \|\vec{r}\| = 1\}$, that corresponds to the surface of a unitary sphere. The dynamic of the population follows a classical attractor network form [5], that is governed by

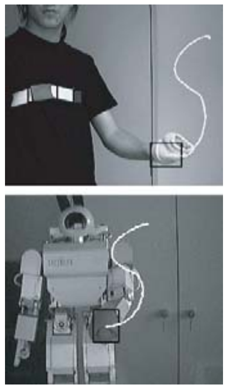
$$\tau \dot{u}_{\vec{r}} = -u_{\vec{r}} + \oint_{\Gamma} w_{\vec{r}\,' \to \vec{r}} \, f(u_{\vec{r}\,'}) \, \mathrm{d}\vec{r}\,' + \vec{v} \cdot \vec{r} + h \quad \text{with} \quad w_{\vec{r}\,' \to \vec{r}} = \gamma \left(\vec{r}\,' \cdot \vec{r} \right)$$

where γ is a system parameter, $w_{\vec{r}\,' \to \vec{r}}$ the lateral weights, $u_{\vec{r}}$ the neuron's membrane potential with preferred direction $\vec{r}$, and $f(u_{\vec{r}})$ its firing activity. In addition, $\vec{v}$ is a vectorial input, and h an homogeneous input applied to the whole population. Such network has been shown to exhibit several properties such as gain modulation [5]. We extended this network [6] to encode independently two separate quantities, namely the direction of $\vec{v}$ and the amplitude h, regardless of the intensity of the directional input. In other words, given a vector $\vec{v}$ and a scalar h, the output population vector will tend toward $h \frac{\vec{v}}{\|\vec{v}\|}$. Thus, following classical linear algebra, several instances of this network could form a vectorial basis. Moreover, assemblies composed of these building blocks result in gain fields, that have been shown to be a neural substrate where multiple sources of information can combine and produce various kind of non-linear transformations [1]. Finally, in order to perform the required transformation, we built a network that is illustrated in Figure 2. It implements the following frames of reference transformation formula

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

where, as shown on Figure 1, $\vec{v}$ and $\vec{v}_T$ are the observed hand and the demonstrator location, $\vec{e}\,'_{i\in\{1..3\}}$ the principal axes, and $\vec{v}\,'$ the final hand location in body-centered FR. The architecture mainly consists in three parallel gain fields that produce, as concurrent projecting outputs, the dot product between the input vector and the principal axes. As a result, the network output converges toward $\vec{v}\,'$, that is the transformed hand location.

Robotic Implementation



We implemented this network in a humanoid robot. The visual inputs are given by a color-based stereo vision system that allows the simultaneous 3D tracking of a human demonstrator's hand, body and principal axes. These information are fed into the neural network that compute the target location in the demonstrator's body centered reference frame. It is then directly mapped to the robot egocentric frame of reference, so that it can immediately imitate the human's hand trajectory using a classical inverse kinematic algorithm. As illustrated here on the right, the robot is able to mimic a gesture shown by a human demonstrator that is not perfectly facing the cameras.

Conclusion

The model described here, provides an example of a plausible neural mechanism for the representation of others in an egocentric frame of reference. Its present implementation has focused on a body-centered frame of reference transformation, as could be found in STS. It is, however, quite general, and could also be applied to object or goal-centered representations. Indeed, these representations are crucial for generalization abilities, as for instance in goal-directed actions, and therefore in 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

1. S. Deneve, P. E. Latham, and A. Pouget. Reading population codes: a neural implementation of ideal observers. *Nat. Neuro.*, 2(8):740–745, 1999.
2. D. Marr. Vision. *Cambridge MA, MIT Press*, 1982.
3. D. I. Perrett, M. Harries, A.J. Mistlin, and A.J. Chitty. Three stages in the classification of body movements by visual neurons. In H.B. Barlow et al. editor, *Images and Understanding*, 94–107. Cambridge University Press, 1989.
4. G. Rizzolatti, L. Fogassi, and V. Gallese. Neurophysiological mechanisms underlying the understanding of actions. *Nat. Rev. Neuro.*, 661–670, 2001.
5. E. Salinas and L.F. Abbott. A model of multiplicative neural responses in parietal cortex. *Proc. Nat. Acad. Sci. USA*, 93:11956–11961, 1996.
6. E. Sauser and A. Billard. Three dimensional frames of references transformations using recurrent populations of neurons. *Neurocomp., In press*, 2005.