



Reaching and Grasping: what we can learn from psychology and robotics

Philippe Gaussier, Alexandre Pitti

► To cite this version:

Philippe Gaussier, Alexandre Pitti. Reaching and Grasping: what we can learn from psychology and robotics. 2017. hal-01609659

HAL Id: hal-01609659

<https://hal.science/hal-01609659>

Preprint submitted on 9 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reaching and Grasping: what we can learn from psychology and robotics

Philippe Gaussier^{1+*} and Alexandre Pitti^{1+*}

¹ETIS, ENSEA, University of Cergy-Pontoise, CNRS, France

*name.surname@u-cergy.fr

+these authors contributed equally to this work

ABSTRACT

Introduction

Grasping an object is an elementary behavior that looks easy for an adult. Yet, grasping an object is still a challenging topic in robotics. Classical approaches consider a sequence of sub tasks ranging from object recognition and localization, the planning of the trajectory to reach this object with the correct orientation and finally the control of the arm movements to grasp securely the object. If this approach has proved to be efficient in simple cases such as reaching a cup on an empty table, a lot of problems remain when the object or the environment is complex. Following,^{1,2} a lot of works have shown the interdependence and even the overlapping between the brain structures involved for real and imagined hand movements. Moreover, some recent works show that the grasping trajectory of an object is impacted by the social environment.³⁻⁶ Grasping an object in order to give it to somebody else is performed differently than picking it up to place it somewhere else. Even if the deposit place is the same, the global trajectory and especially the preparatory movement to pick the object is different. Hence, our brain is perhaps not planning the grasping as a sequence of elementary and independent subtasks. Such studies on the precise modeling of the human arm control can be found in,^{7,8} but is out of the scope of the present paper.

Our aim in this review is to present some underlying neuro-computational mechanisms that can serve for such multimodal integration in infants in order to reach and grasp objects. We will show first how reaching a place or an object can be seen as learning sensory-motor attractors. We propose in our framework that recognizing the object we try to grasp is not performed before the grasping but is the result of the whole grasping dynamics.⁹ The grasping “procedure” can be seen as the building of an attraction basin around the desired object. To succeed such a task, we will insist on the importance of using the same kind of topological coding from sensory areas to motor areas. We will show also several neural mechanisms found important for multimodal integration, reaching and grasping, as well as the importance of sighted touch for anticipatory movements.

Reaching as an emergent property of visuo-motor coordination

According to Bremner, the way infants perceive the space around them (i.e., infants’ spatial representation) relies on two different mechanisms that mature separately during the first year.¹⁰ One is ego-centric and achieves a spatial correspondence of the default body parts and the other is allocentric and helps to localize dynamically the position of the limbs. For instance, infants find hard to locate their own hands when their arms are crossed.

It is assumed that the infant brain exploits strongly Hebbian learning to acquire those spatial maps.¹¹⁻¹³ Hebbian learning can provide to neurons the ability to correlate the contingent events, even across multiple modalities. For instance, infants even at birth can connect the modulation of a sound (from low pitch to high pitch) with the modulation of a light intensity (from low brightness to high luminosity¹⁴), they can relate the contingency between their legs motion and their visual location displayed on a TV monitor^{15,16} or between the texture of a pacifier on mouth and its visual shape.¹⁷

Tenants of the embodied system approach suggest that infants learn the regularities in the structure of sensorimotor information to shape the neural activity in the brain. In this line, sensorimotor neurons become sensitive to the geometrical features of objects in space (their position, their orientation) relative to their bodily gestures and postures; what we call sensorimotor primitives. In return, these neurons would guide and constrain the baby behaviours toward its affordances on the environment; e.g., to orient oneself in the environment or to grasp one object in the correct orientation relative to the body.

All-in-all, we propose that the contingency found across the body signals during sensory-motor exploration –, coming from the arms muscles spindles, the joint angles from the shoulder-elbow-wrist system, the tactile information from the hand, the

sound of a noisy object in the hand and the eye's vision cells as well as their orientation,— may organize the cortical memory into a map of "reachable regions" cells via Hebbian learning. Following this, we can propose that the parietal cortex —, which is acknowledged as the center for multi-modal integration and body representation,— learns body-place associations by creating parieto-motor "reaching cells" for manipulation tasks¹⁸ (see Fig. 1 a and b respectively).

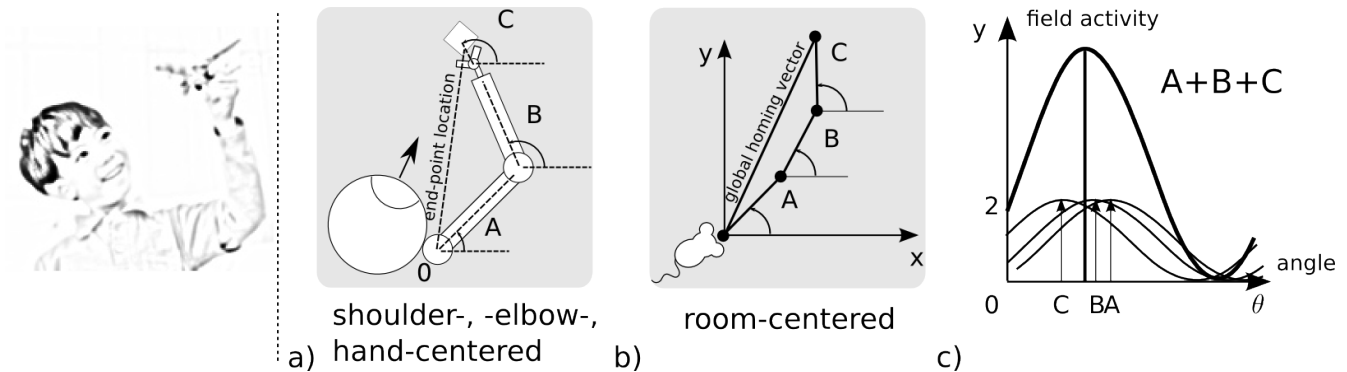


Figure 1. Child holding a toy in his hand. Reaching, grasping and holding objects are non trivial tasks. In a), they require to coordinate one's own actions relative to the objects' relative spatial location and orientation in eye-, hand-, or body-centered reference frames. In b), Reaching and grasping with the hand are similar in essence to a navigation task in a third-person perspective toward a place.¹⁹ In c), grasping can be viewed as performing homing: integrating information by aligning different reference frames through time.

We suppose that the learning of *reach cells* is presumably done by discovering the correct binding that relates each stimulus in its correct reference frame with others in different reference frames; for example this would help for locating the hand in eye coordinates (self-recognition), one object in hand-centered coordinates (relative distance) or one object in shoulder-centered coordinates (body-centered distance).

Those reach cells require to combine contextually or conditionally different sensory signals to represent one location, even if they change dynamically; e.g., if an object is placed at a relative distance of the arm independent of the arm location. For instance, tactile information is dependent on the fingers in the hands, but also on the wrist orientation and arm location, see Fig. 2 a) and b). The visual information is dependent of the eye angle in its orbit (eye-centered) *and* to the head orientation (body-centered), for which the later is only valid for auditory signals.

Considering the neural mechanisms involved in the construction of reach cells, visuomotor neurons have been found to encode the Preferential visual movement Direction (PD) in extrinsic coordinates (likely shoulder-centered) in 3D space.^{20,21} Kakei and colleagues for instance found some muscle-like neurons based on the composition of cosine functions tuned to visual PD and modulated monotonically with the limb posture, see the neuron's activity in Fig. 2 a). When the wrist rotates, the extrinsic neurons discharge proportionally to the preferred visual direction in order to compensate the muscles displacement.²² Other researchers like Scherberger and colleagues found some hand-centered cells whose activity depends both on the object features like the size and the orientation and on the grip type (pronation or supination),^{23,24} see Fig. 2 b). These reach neurons are multimodal and context-dependent. They compute *where* an object is located with respect to the hand position and *how* it is oriented with respect to the hand posture. They are called also in the literature Gain-Field neurons (GF), because their gain-level or amplitude is varying depending on multiple inputs to represent the current object-body relation^{25,26} as shown in Fig. 2.

Hand Orientation based on the Shape of the Object

The gain-field modulation mechanism describes the phenomenon where the motor and the sensor signals (resp. A and B) mutually influence the amplitude activity of their afferent parietal neurons (resp. C).^{27,28} As an example, one possible modulation is the multiplicative effect across the unimodal neurons, A and B , so that the afferent multimodal neuron C possesses the activity $C = A \times B$. Following this, these neurons encode stimulus location simultaneously in more than one reference frame using basis function or 'gain fields'.^{29,30} For instance, there is a non-linear dependency on eye position for certain visual neurons in posterior parietal neurons (PPNs) whose reference frame is centered on the head, whilst others are found to be influenced more by the coding of somatic information into hand/arm-centered reference frame. The same is found for audio-visual signals integration.^{31,32}

Gain modulation contributes therefore as a major computational mechanism for coordinates transformation and for the compensating of distortions caused by movements.²⁶ Its role is even broader as PPNs are found also important for reaching targets, goal-directed movements³³ and even for intentional acts.³⁴ Pouget and Deneve suggest that the parietal neurons behave

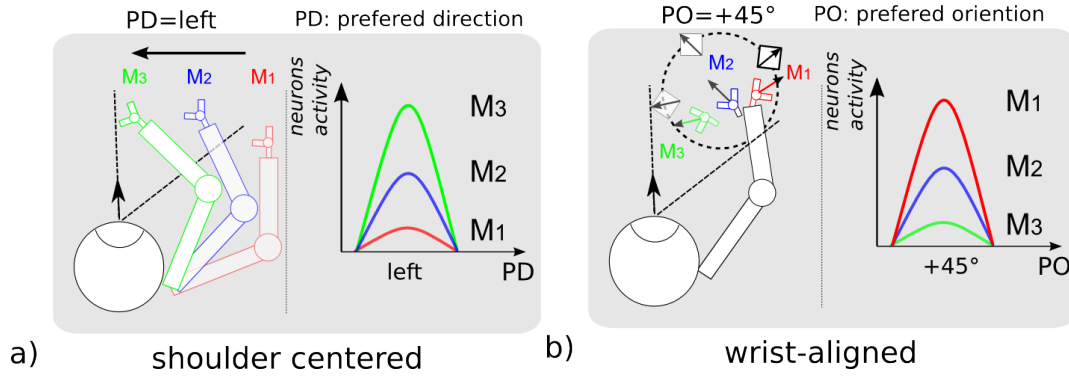


Figure 2. Principle of gain-modulation in parieto-motor neurons (gain-field neurons) for the arm reaching and orientation. Gain-field neurons combine visuo-motor features to encode one preferred direction (PD) or one preferred orientation (PO). In a), exemplar activity of one GF neuron with visual PD *and* when the arm moves on the left. In b), exemplar activity of one GF neuron with some visual PO *and* when the wrist is on pronation.

as a population of basis functions that are continuously adapting their dynamics to the current coordinate frame relative to the task.^{29,35}

Kakei and colleagues²² proposed a three stages architecture to explain the extrinsic-to-intrinsic transformation necessary for hand reaching using visual directional cells. It is inspired from the model of directional motor cells discovered by Georgopoulos.^{36,37}

At the first stage, extrinsic-like neurons A and B encode some preferred directions (PD) of the movement orientation toward θ_A and θ_B respectively for the neuron A and B in the form of cosine-like functions: $A = \cos(\theta - \theta_A)$, $B = \cos(\theta - \theta_B)$.

At a second stage, the gain levels λ_A and λ_B receive incoming signals from various modalities (e.g., visual, somatopic, proprioceptive or auditory) and modulate the amplitude of the extrinsic neurons proportionally to it, so that we have the gain-modulated extrinsic neurons A' and B' equal to: $A' = \lambda_A \cos(\theta - \theta_A)$, $B' = \lambda_B \cos(\theta - \theta_B)$.

At the third stage, the gain-modulated extrinsic-like neurons A' and B' are combined linearly to produce the generation of muscle-like neurons C and the encoding of the motor angle into $C = A' + B'$.

In this motor scheme, the input activity of the gain-levels λ_A and λ_B are connected to the preferred direction in the visual space toward objects so that the motor responses of these neurons are linked to the goal-directed actions toward these objects. The muscle response C , will be either (1) the shoulder-elbow-hand articular system in the reaching task, (2) the wrist muscle in the close reaching posture of the hand, or (3) the wrist muscle in the close wrist/object relative orientation task.³⁸

Emergent control of different joints with one neural field

When turning our eyes, our head, or our arm in a given direction, some specific subpopulations of neurons are activated.³⁹ These neurons have a tuning curve activity that looks like a bell curve according to their preferred direction. Following the brain anatomical projection from the primary visual area to the parietal cortex¹ and the motor cortical area, one simple solution is to suppose the topological organization of the information is preserved (or rebuilt). Electrophysiological recording of "motor" neurons activated by somatosensory and visual stimulation⁴⁰ seem to confirm this hypothesis. Using similar coding from the sensory cortical areas to the motor cortical areas can hugely simplify the control problems. For instance, if we suppose the azimuth of a target in the retinal field is projected onto a larger field taking into account the eye orientation and the head orientation thanks to some kind of neuronal shifting mechanisms⁴¹ the same kind of neural field can be used to code the azimuth of a target in the body space and to control the direction of the movement to reach it. In the dynamical field theory, this population coding is modeled thanks to a field of neurons topologically organized. Recurrent excitatory and inhibitory connections following a Mexican hat shape allow to maintain the goal position as a stable attractor for intermittently detected targets⁴² and to control the body and arm movements. Once the target has been located in the image, 2 one-dimensional fields can be used to minimize the angular position between the target and the arm end effector. One interesting readout of the neural field activity is to consider the spatial derivative of the field as the speed set-point.⁴³ Starting from an aligned body configuration (with the eye, head, body and arm in the same direction), the system will read the same speed set-point for all the different joints. Yet, because of the inertia differences between the different parts of the body, after one step the different joint will end at different azimuth. Hence for the second time step their speed command will become different inducing the feeling that the different degrees of freedom are controlled by different dynamical equations while all their activities is the result of a single one dimensional field. Generalizing this principle both for the azimuth and the site, it is possible to model a smooth trajectory

to reach an arbitrary target with the eyes and the head turning first, followed by the arm and next the whole body reorienting in front of the target (see fig. 3). A key point is clearly the importance of the low level control and the use of proprioceptive information (efferent copy of the action) to perform a continuous planning as opposed to a preplanned trajectory.

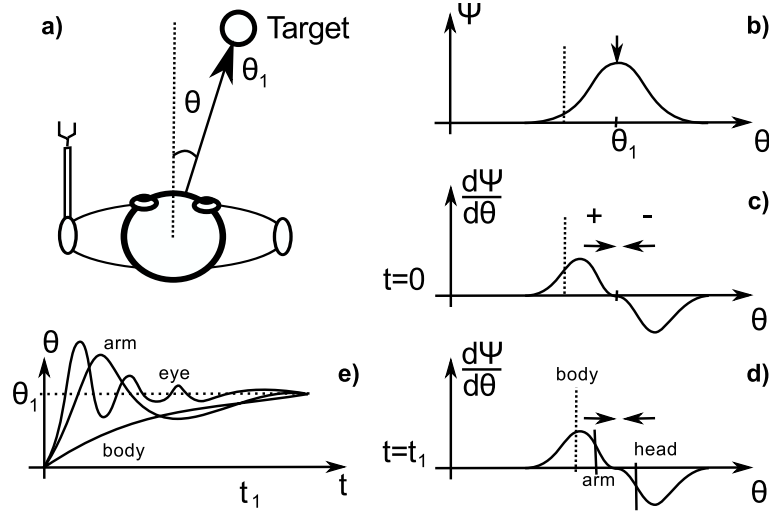


Figure 3. Different apparent dynamics from the same neural field. The body, arm and eye appear to have 3 different dynamics according to their inertia.

The proposed strategy is working correctly only if the different degrees of freedoms are well aligned with vision i.e. if they are either purely in the vertical or horizontal plane. For more complex cases, the visuo-motor transformation must be learned. It can be done using some random motor exploration allowing associating visual position with a specific configuration of the different joints. When the different critical positions have been learned, a neural network can easily interpolate for intermediate positions⁴⁴⁻⁴⁶ allowing the reaching of the visual position. It can be used for gesture imitation or for object grasping.

Building an attractor basin for reaching or grasping

In real life condition, grasping an object is difficult because learning cannot be perfect. If a known object is presented under an unusual position it can be not trivial to decide about the best way to grasp it. How to orient the hand, how large to open the hand... The problem can become even more challenging for novel objects. Where is the right place and orientation to grasp it? Generalization capabilities are a key feature to allow grasping. In the architectures we develop, we suppose a pattern (here an object) can never be perfectly recognized (perception is always ambiguous). Hence, recognizing an object cannot be seen as a all or nothing task.

For instance, if we have learned from three places in the vicinity of a goal location what is the action to reach that place, then it is possible to reach that location from any starting place if the environment remains the same (see fig. 4). To succeed, the only constraint is to suppose that place or state recognition decreases with the distance to the learned position. In that case, in a two-dimensions environment, learning 3 places around the goal and taking the action associated to the better recognized place is sufficient to reach the goal location whatever the starting position. The only constraint is that the starting place belongs to the same visual environment (i.e the starting place is in the neighborhood of the attractor place). The learning creates an attraction basin and the robot is then nothing more than a ball going to the center of the valley defined by the different state/action learned associations. Hence, it is not important to "recognize" one state but to decide the winner state using a competition procedure. At the opposite to classical pattern recognition, we will not suppose that recognizing a shape is performed when the recognition level is higher than a given threshold. Recognizing a shape is related to the sensorimotor sequence of actions allowing reaching that state. It can be performed even if the different states are badly recognized. The only constraint is that the recognition decreases when moving away from a learned state. The same way reaching an object is equivalent to some extent to recognizing this object and can be performed using the same kind of architecture (see fig. 5).

When grasping an object, the mechanical constraints induce some stable proprioceptive-motor states. From the visual information (for instance some Gabor signatures) it is possible to predict a field where the maximum activity corresponds to one of the orientations to grasp the object. As a result, learning the vision of the object will induce a field activity with as many maxima as the number and positions of the possible grasping (see fig. 6).

Using this field as the goal for reaching will allow the arm to move in the direction of the nearest "good" grasping location on the object (moving according to the derivative of the field). It is not clear if this field has to be explicit in our brain (what we

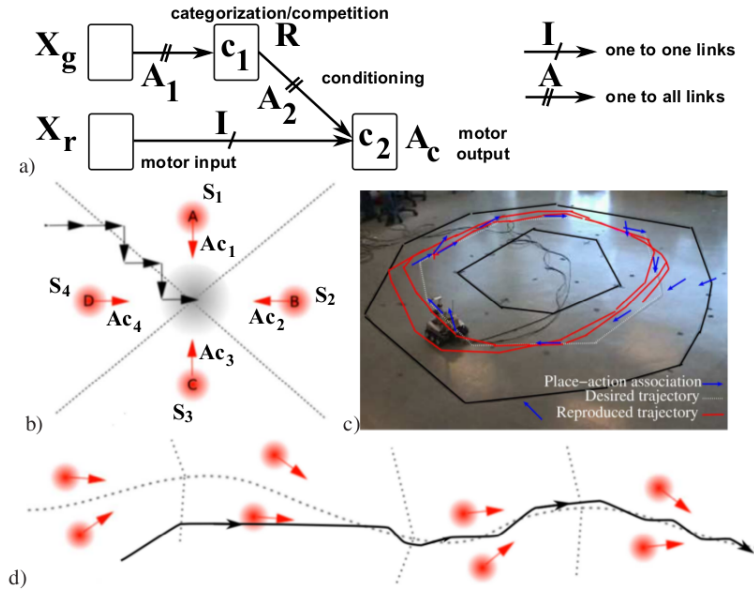


Figure 4. a) PerAc model. b-d) Examples of built dynamics in 2D spaces. b) Fixed point attractor. c) Limit cycle in the case of navigation experiment. d) Trajectory following. In b) and d), the gray dotted lines are the Voronoi boundaries. The plain black line is a trajectory sample.

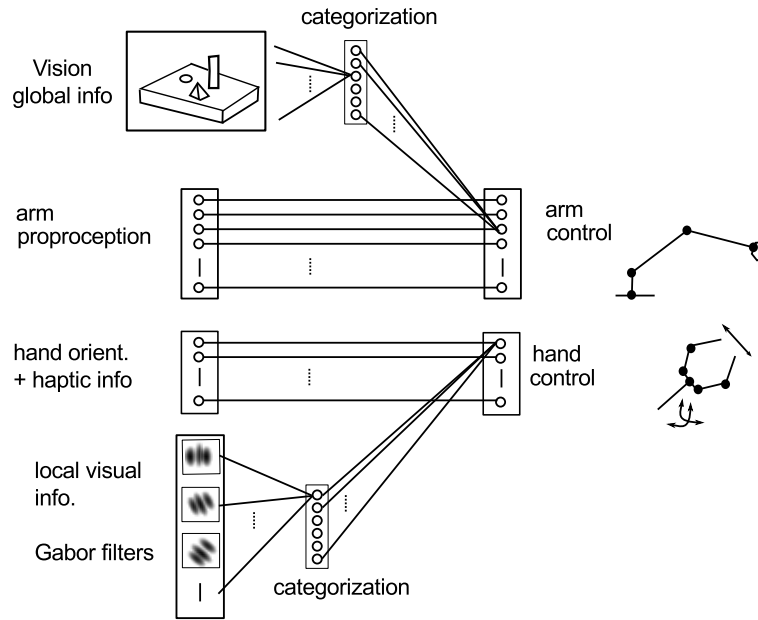


Figure 5. PerAc architecture for the visual control of object grasping.

consider as a field could be the projection on the action space of the high dimension sensory-motor connectivity). Moreover, the learning of two views of the same objects and their associated actions can be sufficient to induce a competition allowing the grasping and can be interpreted as the result of a virtual neural field.⁴⁶ Hence the attractor could be completely emergent and not formally endowed in a neural field (the neural field being an abstract way to represent the attractor reached by the arm/hand couple).

It is also clear that recognizing an object involves not only cortical areas but also subcortical areas. Once the object has been grasped, it is possible to use the dynamical expectations of the arm movement as a way to recognize some important feature of an object such as its weight. As a matter of fact, lifting an object will induce different trajectories according to its mass and to the force used. Knowing (and deciding) the force profile, it is possible to predict the timing of the arm proprioception. Then a

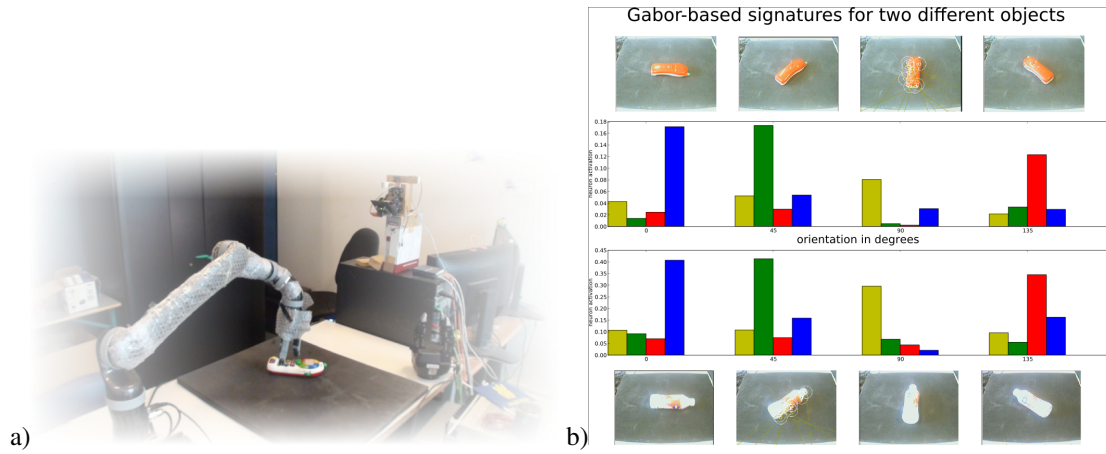


Figure 6. a) A simple robotic setup for object grasping. b) Predicting object orientation for grasping according to the result of Gabor filters on the object image.

simple difference between the temporal predictions according to the force used and the recognition of the object is sufficient to decide if the object lifted was the correct object or not.⁴⁷ In our case, we are using a simple model of the cerebellum to predict the temporal dynamics of a given force command in a given visual context (i.e. the recognition of a given object).^{48,49} It allows to predict if there is a mismatch between the object visual appearance and the object inertia.

Multimodal Strategies for Reaching & Grasping

Perceiving objects in space is one of the first tasks babies have to deal with during infancy. It is a rather difficult problem since infants have to represent one object with multiple sensory modalities (vision, sound, touch) encoded in different reference frames (e.g., eye-centered, head-centered, or hand-centered). This curse of dimensionality corresponds to the so-called binding problem across the modalities and requires at least to construct one amodal and unified reference frame relying on coordinates transformation between the senses *or* to form a network of partially inter-connected reference frames to represent one stimulus in each modality.⁵⁰

Touch is often neglected when we consider reaching and grasping tasks. However, these actions are goal-directed toward manipulating objects. Although tactile information is acquired at the very end of these tasks, it has an influence on the whole action sequence from the beginning till its end, by *anticipating* the orientation of the hand with respect to the object to be manipulated and touched. The predictive coding done by the somatosensory neurons, combined with other senses (vision, proprioception), may give rise to the emerging capabilities of the so-called mirror neurons, invisible to the eyes like body image,^{51–54} anticipatory touch,^{55,56} action observation,¹³ social cognition⁵⁷ as well as tool-use⁵⁸ and language.⁵⁹

Touching objects permits to discover the physical limits of the body and to construct its spatial representation with the use of vision and proprioception,⁶⁰ see Fig. 7 left graph. This spatial representation is however non-linear, the same tactile stimulus can correspond to different visuo-motor pairs. However, the tactile stimuli can serve to learn the relative locations of these different visuo-motor pairs with respect to the hand motion as in Fig. 1. Over time, the tactile cells can develop extended visual receptive fields with anticipatory effect,^{18,61,62} see Fig. 7 right graph. These effects have been attributed also to the so-called mirror neurons in the premotor area that integrate multisensory information for action purposes.^{53,55,63}

These visuo-tactile receptive fields can be learned from the same gain-modulation mechanism introduced in the previous section. The output motor neuron C , which combines visuo-motor information for hand position/orientation, is here replaced by the tactile neuron T anchored in one location on the skin and relative to the arm posture.

We present some preliminary results that compile experiments done in one robotic arm, the JACO arm from Kinova Robotics, with vision and one artificial skin for self-representation, reaching and grasping motion, see Fig. 8. The framework is the one explained above based on gain-modulation for the modeling of different cells that integrate multimodal information for grasping (hand-object alignment), distance (combining visuo-motor information with one artificial skin on the arm), and reaching (object location in shoulder-, elbow-, hand- center reference frame); resp. a-d).

Fig. 8 a) presents the model proposed by Kakei and described in previous sections to explain the general architecture for extrinsic-to-intrinsic transformations based on gain-field modulation; e.g., for object reaching (visual direction sensitivity of motor cells and hand-centered visual cells), for object-wrist alignment (relative visual orientation of motor cells). Fig. 8 b) displays the neural dynamics of a gain-field neuron sensitive to one object relative orientation to the robot wrist after motor

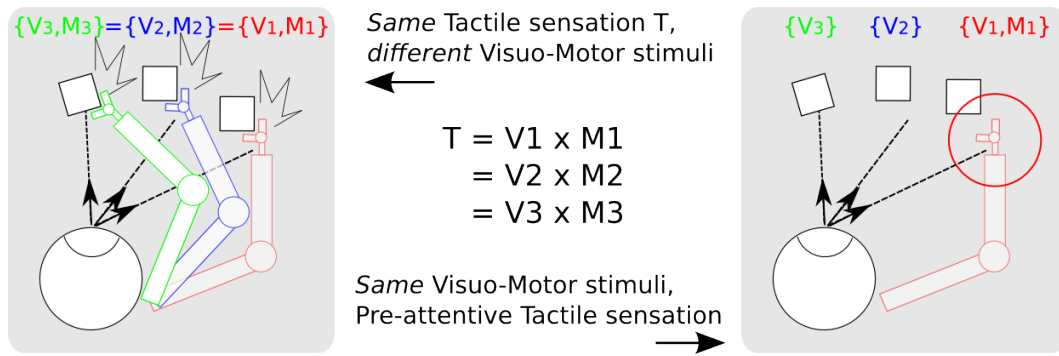


Figure 7. Tactile information is important for calibrating and anticipating spatial locations for reaching and grasping tasks. On the left, one task that reach cells have to resolve is to link the same tactile cell with different visuo-motor pairs. On the right, when spatial locations have been learned by reach cells with the mechanisms of multimodal integration explained above, they can serve for anticipating visually one touch; acquisition of body image.

babbling. The graph shows five wrist motor angles with respect to the global visual orientation of the stick (colormap hsv, resp. 0 to 180 degrees). This graph shows how the neuron amplitude is modulated depending on the wrist orientation (intrinsic coordinates) and to the stick orientation (extrinsic coordinates); this neuron permits the transformation from intrinsic to extrinsic coordinates. The shift in preferred orientation (PO) of the neuron corresponds to the graded contribution of the GF neuron.

Fig. 8 c) reproduces the experience for body-centered representation with a robotic arm and its artificial skin. With a similar neural architecture for intrinsic-to-extrinsic transformation, the tactile input perceived from a salient object touching the skin surface is used to localize at the same time its visual location in the image for different motor positions. After the learning stage, it is possible to reconstruct back the relative visuo-tactile receptive fields even when an object is approaching the skin without touching it. The length of each arrow indicates the activity level of the nearby neuron anchored at the visual position of the skin for a specific orientation and distance of any approaching objects.

The last figure in Fig. 8 d) shows a simulation of a reaching task based on the modeling of visual motor cells in 3D. The network learns the relation between posture and visual direction, $A \times B$, to generate a motor command C in that direction.

These different experiments based on multiple modalities show to some points that one mechanism may serve as a unifying principle for spatial representation in order to interact in the world, for robots, brains and babies.

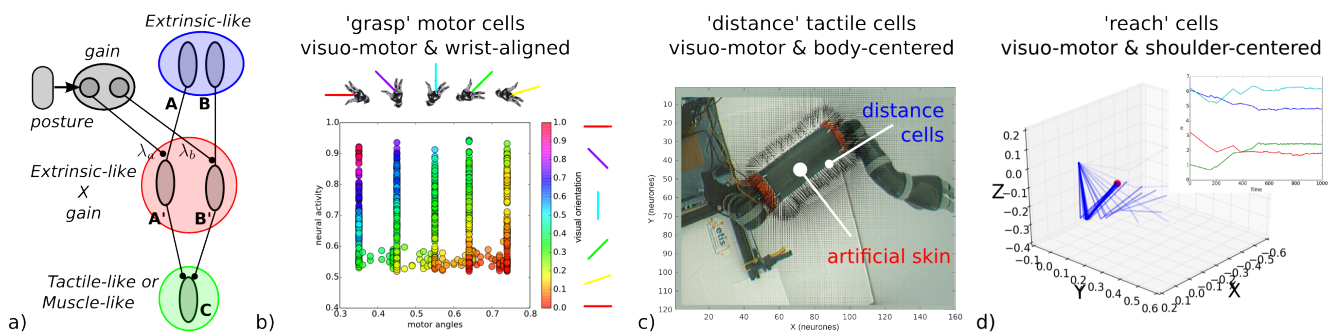


Figure 8. Gain-field neurons for the modeling of grasping, distal and reaching cells. In a), mechanism of gain-modulation for multimodal integration based on multiplying various signals. In b), its modeling for the generation of wrist motor 'grasp' cells aligned to the visual orientation of one object. The neural activity of the neuron is dependent of the wrist orientation and the visual orientation of the object. In c), the activity of 'touch' cells on the artificial skin are super-imposed to integrate distance depending on the visual position on the camera and the arm orientation. In d), reach cells are modeled in this 3D simulation of the arm by summing up cosine/gaussian functions as in a). It is similar to path integration framework explained in Fig. 1 c).

Discussion

The neurons that learn the proper sensory-motor associations of the body may be useful for perceiving someone else actions by simulating them. They will react in a similar way to mirror neurons.^{62,64} They are either activated by the action or by the recognition of the visual features associated with the object to grasp. Hence, gesture imitation and action imitation could

be seen, at first, as the result of the perception ambiguity. Being unable to recognize his/her own body from the body of their parent, could force the baby or the robot to mimic their parent/caretaker gestures as a way to minimize the difference between their own proprioception and the visual position observed hand. Indeed, in this case, the only solution for the system to minimize the discrepancy between vision and proprioception is to move his arm in a way inducing an emergent low level imitation behavior.^{64,65}

For learning to grasp and avoid objects, it is necessary to add emotional mechanisms as a primary conditioning system allowing to associate positive and negative object with grasping and avoidance behaviors. If we suppose a cascade of sensorimotor architectures⁴⁴ allowing first to associate the adult facial expression with some positive or negative values⁶⁶ then, smiling when a baby (or a robot) is grasping an object or doing an angry face can be used as way to trigger object attraction or repulsion for objects which are primarily neutral for the baby or the robot. Hence, social referencing and complex motor behavior can emerge from the same architecture using some second order conditioning mechanisms.^{67,68}

For even more complex tasks involving tool use, we can suppose first that the graspability of an object can be represented by one or several neural fields representing from where the object can be grasped (the higher, the easier to grasp). The aperture and orientation of the hand can be learned. Next, when an object has been grasped, it can modify the field associated to the hand (or the arm extremity) becoming a tool for the external observer. On the neural field associated to the arm extremity, if the object orientation is on the left, an activity bubble on the left will traduce the capability of the object to act on the left. A "T" object will produce both left and right action capabilities inducing two emergent solution when grasping is necessary. Hence, if we suppose the arm is composed of several degrees of freedom in a serial order, their associated bubble of activities can be summed. As a result the maximum of activity on the azimuth and site fields is associated to the global arm direction and the intensity of the maximum to the arm elongation. Grasping or using a tool would be the result of the same reconfigurable neural field associated to the arm end effector and learned from the dynamical sensory-motor exploration of the object or tool.

References

1. Sirigu, A., Duhamel, J.-R., Cohen, L., Pillon, B. *et al.* The mental representation of hand movements after parietal cortex damage. *Science* **273**, 1564 (1996).
2. Gerardin, E. *et al.* Partially overlapping neural networks for real and imagined hand movements. *Cerebral cortex* **10**, 1093–1104 (2000).
3. Becchio, C., Sartori, L., Bulgheroni, M. & Castiello, U. The case of dr. jekyll and mr. hyde: a kinematic study on social intention. *Consciousness and cognition* **17**, 557–564 (2008).
4. Becchio, C., Sartori, L., Bulgheroni, M. & Castiello, U. Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition* **106**, 894–912 (2008).
5. Lewkowicz, D., Delevoye-Turrell, Y., Bailly, D., Andry, P. & Gaussier, P. Reading motor intention through mental imagery. *Adaptive Behavior* 1059712313501347 (2013).
6. Quesque, F., Lewkowicz, D., Delevoye-Turrell, Y. N. & Coello, Y. Effects of social intention on movement kinematics in cooperative actions. *Frontiers in neurorobotics* **7**, 14 (2013).
7. Bullock, D., Grossberg, S. & Guenther, F. H. A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience* **5**, 408–435 (1993).
8. Bullock, D. & Grossberg, S. Vite and flete: Neural modules for trajectory formation and postural control. *Advances in Psychology* **62**, 253–297 (1989).
9. Kelso, J. S. *Dynamic patterns: The self-organization of brain and behavior* (MIT press, 1997).
10. Bremner, A., Holmes, N. & Spence, C. Infants lost in (peripersonal) space? *Trends in Cognitive Sciences* **12**, 298–305 (2008).
11. Del Giudice, M., Manera, V. & Keyers, C. Programmed to learn? the ontogeny of mirror neurons. *Developmental Science* **12**, 350–363 (2009).
12. Heyes, C. Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews* **30**, 575–583 (2010).
13. Keyers, C., Perrett, D. & Gazzola, V. Hebbian learning is about contingency, not contiguity, and explains the emergence of predictive mirror neurons. *Behavioral and Brain Sciences* **37**, 205–206 (2014).
14. Lewkowicz, D. & Turkewitz, G. Cross-modal equivalence in early infancy: auditory-visual intensity matching. *Dev. Psychol.* **16**, 597–607 (1980).
15. Rochat, P. Self-perception and action in infancy. *Exp. Brain Res.* **123**, 102–109 (1998).

16. Shimada, S., Hiraki, K. & Oda, I. The parietal role in the sense of self-ownership with temporal discrepancy between visual and proprioceptive feedbacks. *NeuroImage* **24**, 1225–1232 (2005).
17. Meltzoff, A. & Borton, R. Intermodal matching by human neonates. *Nature* **282**, 403–404 (1979).
18. Graziano, M. & Cooke, D. Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia* **44**, 845–859 (2006).
19. Gaussier, P. *et al.* A model of grid cells involving extra hippocampal path integration, and the hippocampal loop. *Journal of integrative neuroscience* **6**, 447–476 (2007).
20. Blohm, G., Khan, A. & Crawford, J. Spatial transformations for eye–hand coordination. *New Encyclopedia of Neuroscience* **9**, 203–211 (2008).
21. Blohm, G. & Crawford, J. Computations for geometrically accurate visually guided reaching in 3-d space. *Journal of Vision, in special issue 'Sensorimotor processing of goal-directed movements'* **7**, 4, 1–22 (2012).
22. Kakei, S., Hoffman, D. & Strick, P. Sensorimotor transformations in cortical motor areas. *Neuroscience Research* **46**, 1–10 (2003).
23. Gail, A., Klaes, C. & Westendorff, S. Implementation of spatial transformation rules for goal-directed reaching via gain modulation in monkey parietal and premotor cortex. *Journal of Neuroscience* **29**, 9490–9499 (2009).
24. Fluet, M., Baumann, M. & Scherberger, H. Context-specific grasp movement representation in macaque ventral premotor cortex. *Journal of Neuroscience* **30**, , 15175–1518 (2010).
25. Salinas, E. & Thier, P. Gain modulation: A major computational principle of the central nervous system. *Neuron* **27**, 15—21 (2000).
26. Salinas, E. & Sejnowski, T. J. Gain modulation in the central nervous system: Where behavior, neurophysiology and computation meet. *The Neuroscientist* **7**, 430–440 (2001).
27. Andersen, R. & Mountcastle, V. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neuroscience* **3**, 532–548 (1983).
28. Andersen, R., Essick, G. & Siegel, R. Encoding of spatial location by posterior parietal neurons. *Science* **230**, 450–458 (1985).
29. Pouget, A. & Snyder, L. Spatial transformations in the parietal cortex using basis functions. *J. of Cog. Neuro.* **3**, 1192–1198 (1997).
30. Pouget, A. & Snyder, L. Computational approaches to sensorimotor transformations. *Nature Neuroscience* **3**, 1192–1198 (2000).
31. Deneve, S. & Pouget, A. Bayesian multisensory integration and cross-modal spatial links. *J. of Phys.–Paris* **98**, 249–258 (2004).
32. Pitti, A., Blanchard, A., Cardinaux, M. & Gaussier, P. Gain-field modulation mechanism in multimodal networks for spatial perception. *12th IEEE-RAS International Conference on Humanoid Robots Nov.29-Dec.1, 2012. Business Innovation Center Osaka, Japan* 297–302 (2012).
33. Chang, S., Papadimitriou, C. & Snyder, L. Using a compound gain field to compute a reach plan. *Neuron* 744–755 (2009).
34. Cui, H. & Andersen, R. Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* **56**, 552–559 (2009).
35. Deneve, S., Pouget, A. & Duhamel, J. A computational perspective on the neural basis of multisensory spatial representations. *Nature Rev. Neurosciences* **98**, 741–747 (2002).
36. Georgopoulos, A., Kalaska, J., Caminiti, R. & Massey, J. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
37. A.P., G., Merchant, H., Naselaris, T. & B., A. Mapping of the preferred direction in the motor cortex. *Proc Natl Acad Sci USA*. **104**, 11068–72 (2007).
38. Mahe, S., Braud, R., Gaussier, P., Quoy, M. & Pitti, A. Exploiting the gain-modulation mechanism in parieto-motor neurons: Application to visuomotor transformations and embodied simulation. *Neural Networks* **62**, 102–111 (2015).
39. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *The Journal of Neuroscience* **2**, 1527–1537 (1982).
40. Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. Visuomotor neurons: Ambiguity of the discharge or ‘motor’ perception? *International journal of psychophysiology* **35**, 165–177 (2000).

41. Koch, C. & Ullman, S. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, 115–141 (Springer, 1987).
42. Amari, S.-i. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics* **27**, 77–87 (1977).
43. Schöner, G., Dose, M. & Engels, C. Dynamics of behavior: Theory and applications for autonomous robot architectures. *Robotics and autonomous systems* **16**, 213–245 (1995).
44. Gaussier, P. & Zrehen, S. Perac: A neural architecture to control artificial animals. *Robotics and Autonomous Systems* **16**, 291–320 (1995).
45. Burnod, Y. *et al.* Parieto-frontal coding of reaching: an integrated framework. *Experimental brain research* **129**, 325–346 (1999).
46. De Rengervé, A., Andry, P. & Gaussier, P. Online learning and control of attraction basins for the development of sensorimotor control strategies. *Biological cybernetics* **109**, 255–274 (2015).
47. Razakarivony, S., Gaussier, P. & Oueddou, F. B. From force control and sensory-motor informations to mass discrimination. In *International Conference on Simulation of Adaptive Behavior*, 221–231 (Springer, 2010).
48. Blakemore, S.-J., Frith, C. D. & Wolpert, D. M. The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport* **12**, 1879–1884 (2001).
49. Jenmalm, P., Schmitz, C., Forssberg, H. & Ehrsson, H. H. Lighter or heavier than predicted: neural correlates of corrective mechanisms during erroneously programmed lifts. *The Journal of neuroscience* **26**, 9015–9021 (2006).
50. Heed, T., Buchholz, V., Engel, A. & Roder, B. Tactile remapping: from coordinate transformation to integration in sensorimotor processing. *Trends in Cogn. Sci.* **19**, 251–258 (2015).
51. Maravita, A., Spence, C. & Driver, J. Multisensory integration and the body schema: Close to hand and within reach. *Current Biology* **13**, R531–R539 (2003).
52. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev. Neuroscience* **27**, 169–192 (2004).
53. Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P. & Casile, A. Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. *Science* **324**, 403–406 (2009).
54. Pitti, A., Kuniyoshi, Y., Quoy, M. & Gaussier, P. Modeling the minimal newborn’s intersubjective mind: The visuotopic-somatotopic alignment hypothesis in the superior colliculus. *PLoS ONE* **8**, e69474 (2013).
55. Keysers, C. Demystifying social cognition: a hebbian perspective. *Trends in Cognitive Sciences* **8**, 501–507 (2004).
56. Pitti, A., Alirezai, H. & Kuniyoshi, Y. Mirroring maps and action representation through embodied interactions. *Connectionist Models of Behaviour and Cognition II, World Scientific* 27–37 (2008).
57. Keysers, C., Kaas, J. & Gazzola, V. Somatosensation in social perception. *Nature Reviews Neuroscience* 427–428 (2010).
58. Maravita, A. & Iriki, A. Tools for the body (schema). *Trends in Cognitive Sciences* **8**, 79–96 (2004).
59. Rizzolatti, G. & Arbib, A. Language within our grasp. *Trends in Neuroscience* **21**, 188–194 (1998).
60. Pitti, A., Pugach, G., Gaussier, P. & Shimada, S. Spatio-temporal tolerance of visuo-tactile illusions in artificial skin by recurrent neural network with spike-timing-dependent plasticity. *Scientific Reports* to appear (2017).
61. Graziano, M., Hu, X. & Cooke, D. Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology* **77**, 2268–2292 (1997).
62. Pitti, A., Alirezai, H. & Kuniyoshi, Y. Cross-modal and scale-free action representations through enaction. *Neural Networks* **22**, 144–154 (2009).
63. Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. The space around us. *Science* **277**, 190–191 (1997).
64. Gaussier, P., Moga, S., Quoy, M. & Banquet, J.-P. From perception-action loops to imitation processes: A bottom-up approach of learning by imitation. *Applied Artificial Intelligence* **12**, 701–727 (1998).
65. Andry, P., Gaussier, P., Moga, S., Banquet, J.-P. & Nadel, J. Learning and communication via imitation: An autonomous robot perspective. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* **31**, 431–442 (2001).
66. Boucenna, S., Gaussier, P., Andry, P. & Hafemeister, L. Imitation as a communication tool for online facial expression learning and recognition. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 5323–5328 (IEEE, 2010).

67. Karaouzene, A., Gaussier, P. & Vidal, D. A robot to study the development of artwork appreciation through social interactions. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, 1–7 (IEEE, 2013).
68. Boucenna, S., Gaussier, P. & Hafemeister, L. Development of first social referencing skills: Emotional interaction as a way to regulate robot behavior. *IEEE Transactions on Autonomous Mental Development* **6**, 42–55 (2014).

Acknowledgments

We thank Ganna Pugach, Sylvain Mahé and Kemo Adrian for fruitful work on the experimental setups. This work was partially funded by EQUIPEX-ROBOTEX (CNRS), ANR DIRAC, chaire d'excellence CNRS-UCP and project Labex MME-DII (ANR11-LBX-0023-01).