2009 Special Issue

Cross-modal and scale-free action representations through enaction

Alex Pitti [a,*], Hassan Alirezaei [b], Yasuo Kuniyoshi [a,b]

[a] *ERATO Synergistic Intelligence Project, JST, The University of Tokyo, 113-8656 Tokyo, Japan*
[b] *ISI Laboratory, Department of Mechano-Informatics, The University of Tokyo, 113-8656 Tokyo, Japan*

ARTICLE INFO

ABSTRACT

Embodied action representation and action understanding are the first steps to understand what it means to communicate. We present a biologically plausible mechanism to the representation and the recognition of actions in a neural network with spiking neurons based on the learning mechanism of spike-timing-dependent plasticity (STDP). We show how grasping is represented through the multi-modal integration between the vision and tactile maps across multiple temporal scales. The network evolves into a small-world organization with scale-free dynamics promoting efficient inter-modal binding of the neural assemblies with accurate timing. Finally, it acquires the qualitative properties of the mirror neuron system to trigger an observed action performed by someone else.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Before articulating the first words, the development of social cognition starts with non-verbal communication and the understanding of actions performed by others. Perception of movements, gestures and actions of someone else can help us understand (or guess) about his intentions, his desires, and his emotions.

These capacities of non-verbal communication are argued to be formed from the existence of pragmatic representations, generally implicit arose from the intertwining between perception and action within the brain (Hiraki, 2006; Rizzolatti, Fadiga, Fogassi, & Gallese, 1996), they constitute the body schema that activate automatically the motoric representations in the prefrontal and frontal area. It follows that, observing someone else acting, recognizing it, and understanding it may result then from a direct pairing between the visually observed action and our own motoric representation of it. Differently said, the observer mentally "simulates" the action from his own experience of it (Gallese, 2005), leading then to a "resonance entrainment" in his motor system (Rizzolatti & Craighero, 2004; Rizzolatti et al., 1996; Rizzolatti, Fogassi, & Gallese, 2001). This phenomenon, termed mirror neurons – located in the $F5$ area in the pre-motor cortex (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996) – describes the neurons' response to action-related visual stimuli, such as graspable object or action of other individuals.

Of particular importance, mirror neurons show temporal congruence between visual and motor neurons (Oztop, Kawato, & Arbib, 2006): mirror neurons fire with *accurate* timing to both observed and to hidden end-state actions. Visual representations of an observed action are therefore temporally linked to our own motor representations of the same action, a product of associative learning in line with the generalist theories of imitation (Brass & Heyes, 2005; Heyes, 2001). According to them, what facilitates imitation is due to the general organization of motor control rather than a special purpose mechanism dedicated to imitation. Mirror neurons are thus not innate systems, but rather acquired from learned perceptual-motor links. Other evidences from developmental psychology tend to confirm that timing between sensory and motor representation is crucial for babies in order to acquire the significance of one action. For instance, infants identify soon the timing correlations and the sequential order of events; e.g., synchrony and contingency (Prince & Hollich, 2005). Moreover, in interceptive actions such as reaching and grasping, synchrony detection between different sensory and/or motor channels is particularly important for detecting the right timing for contact or that of preparatory actions (Corbetta, Thelen, & Johnson, 2000; Prince & Hollich, 2005). More complex cognitive abilities – e.g., imitation, self-agency and social interaction – may be developed from these newly acquired affordances (Heyes, 2004; Meltzoff & Moore, 1977; Nadel, Prepin, & Okanda, 2005; Rochat, 2003; Zukow-Goldring, 2005). Taken together, these considerations suggest that exploiting the mechanism(s) regulating timing at the neural level can reveal some of the principle(s) behind action representation, cognitive development and social interaction.

At the neural level, the regulation mechanism responsible for the timing delays between the spikes is the one of spike-timing-dependent plasticity termed STDP (cf. Bi and Poo (1998) and also Abbott and Nelson (2000)). Temporal structure of complex

---

* Corresponding author.
E-mail address: alex@jeap.org (A. Pitti).

actions, for instance, are decomposed with millisecond order precision into ordered sequences of neural rules in canonical motor neurons and in the mirror neuron system (Changeux & DeBevoise, 2004; Lestou, Pollick, & Kourtzi, 2008; Rizzolatti et al., 1996). Precisely, information processing in large networks of spiking neurons is performed both in the temporal domain (i.e., time delay between the spikes) and in the spatial domain (i.e., spatial location of the neurons). It is therefore the coherency of the local dynamics among the neural pairs that will (or will not) produce a coherency at the network scale—we mean a functional integration among the different parallel processes in the maps into a dynamical representation of the body in action.

Our main objective is to understand how such global integration in the neural dynamics is produced during physical interactions. How functional connectivity in the network permits the representation of one action from the differentiated processes done in the sensor and the motor maps having a structured multi-modal activity of the neural code. In this paper, we demonstrate how actions are represented at the neural level as accurate spatio-temporal clusters sparsely encoded over distant neural maps ruled by the learning mechanism of STDP. We set up an experiment of grasping, in which the temporal sequence of the action is acquired (or "represented") through the neural interaction between vision and tactile modalities. This functional integration – termed "vertical association" by Brass and Heyes (2005) – between the sensory and motor maps produces the emergent structure of reentrant or mirroring maps, a result of their entanglement due to embodiment. Interestingly, reentry achieves the cross-modal linkage between the tactile and vision maps making the neural system earn the capabilities of associative memory. For instance, inter-modal activation (capacity to trigger one modality from another) and anticipation (capacity to anticipate the next state of the other modality) combining the feature of a coupled forward and inverse model, predicting the sensory consequences of a motor command and transforming a desired sensory state into a motor command that can achieve it (Oztop et al., 2006). Since the network produces inter-modal associations, information may be retrieved back from the activation of another modality. It follows that the observation of one action (i.e., visual information available only) will induce the simulation of the missing modality (i.e., haptic perception). The qualitative property observed in the mirror neuron system.

In the first section, we present the framework employed to design our neural network. Thereinafter, we study how the network acquires appropriate perception–action matching from repeated experiences of seeing and touching permitting to reproduce the qualitative properties of canonical and mirror neurons: firing to executed actions and to observed actions. We then discuss the relevance of our findings to cross-modal binding and to functional integration in the brain. We advance that the neural organization of the mirror neuron system is mediated by the regulatory mechanism of STDP for action representation and action understanding using the same pathways.

2. Framework

In comparison with classical feed-forward neural networks, information processing in recurrent networks of spiking neurons is not based on the statistical modeling of the available data but rather on the parallel processing of the neurons combined in a self-organized fashion (i.e., assembling the relative spatio-temporal coordinations). We define, in this part, the network architecture, the neuron model used in our experiments and the reinforcement mechanism of spike-timing-dependent plasticity (STDP) that regulates the dynamics of the neurons from each other. We then detail the design of the retina model (biologically inspired) used for the visual processing in our experiments to the transforming of intensity-based images into spike trains.

2.1. Spiking neuron model

The neurons are defined with the formal model (temporal derivatives) proposed by Izhikevich [cf. Izhikevich (2003)]:

$$v' = 0.04v^2 + 5v + 140 - u + I$$
$$u' = a(bv - u) \tag{1}$$

with v representing the membrane potential of the neuron in mV and u a membrane recovery variable – v' and u' their respective temporal derivatives. The neurons are externally triggered by the signal I and their dynamics are resetted after any spiking

$$\text{if} \quad v \geq +30 \text{ mV}, \quad \text{then} \begin{cases} v \leftarrow c \\ u \leftarrow u + d. \end{cases} \tag{2}$$

The variables set $\{a, b, c, d\}$ defines the neuron attributes whether excitatory $(a; b) = (0.02; 0.2)$ and $(c; d) = (-65; 8)$, or inhibitory; $(a; b) = (0.02; 0.25)$ and $(c; d) = (-65; 2)$. For further details, see Izhikevich (2003) and Izhikevich, Gally, and Edelman (2004).

2.2. Recurrent neural network architecture

In our experiments, the networks are composed of large ensembles of neural units. The neurons are connected to each other with arbitrarily short- and long-range synaptic connections (up to one hundred synaptic links for each neurons) and with variable time delays between the neurons (arbitrarily defined up to 20 ms). By doing so, information is sparsely coded in the recurrent networks which facilitates the recall of memories from partial cues and allow for denser and more reliable storage (Aoki & Aoyagi, 2007). Without external constraints from the environment, no particular organizational structure is visible at the system level which gives rise to a spontaneous-like activity in its dynamics.

We explain in the following part the details on the mechanism of spike-timing-dependent plasticity (STDP) on which the networks rely on.

2.3. Reinforcement mechanism of spike-timing-dependent plasticity

STDP is the bidirectional adaptation mechanism which dynamically regulates the long-term potentiation (LTP) and long-term depression (LTD) in synaptic plasticity readjusting the synaptic weights to the precise timing interval between the initiating and the targeting neurons (Abbott & Nelson, 2000; Bi & Poo, 1998; Song, Miller, & Abbott, 2000). They are significant mechanisms for both activity-dependent development of neural circuitry and adult memory storage. The time delay $\Delta t = t_{\text{post}} - t_{\text{pre}}$ between the pre-synaptic neuron spiking t_{pre} and the post-synaptic neuron firing t_{post} corresponds to the interval range of activation of their synaptic plasticity and weight adaptation Δc.

$$c_{\text{pre,post}} = c_{\text{pre,post}} + \Delta c \tag{3}$$

$$\Delta c(\Delta t) = \begin{cases} A_+ \exp(\Delta t/\tau_+) & \text{if } \Delta t < 0 \\ -A_- \exp(-\Delta t/\tau_-) & \text{if } \Delta t \geq 0. \end{cases} \tag{4}$$

The synaptic weights decay exponentially depending on the time delay Δt between the pre- and post-synaptic neurons in the interval range $[\tau_-, \tau_+]$ [see Fig. 1(a)]. Each time a post-synaptic neuron fires, its synaptic weights $c_{\text{pre,post}}$ are decreased by A_- (LTD), and each time a synapse receives an action potential, its synaptic weight $c_{\text{pre,post}}$ is incremented by an amount A_+ (LTP). In all our experiments, we set $-A_- = A_+ = 1$ and $\tau_- = \tau_+ = 20$ ms. Over time, pairs of neurons are consolidated and can form long-range clusters of parallel processes [see Fig. 1(b) and Fig. 2], the idea behind polychronization coined by Izhikevich (2006) and Izhikevich et al. (2004) that we present hereinafter.
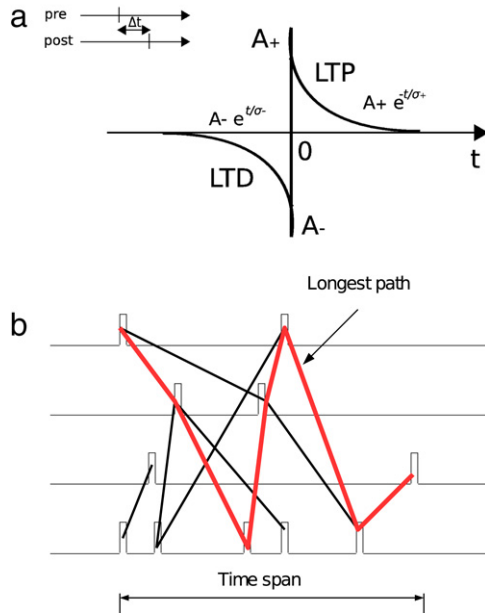
Fig. 1. Mechanism of STDP with $-A_+ = A_- = +1$ and $\tau_- = \tau_+ = 20$ ms. (a) Each time a post-synaptic neuron fires, its synaptic weight is decreased by A_-, and each time a synapse receives an action potential, its synaptic weight is incremented by an amount A_+. (b) Neuronal groups are formed from the dynamical linkage between the neural pairs (hierarchical representations).

Spike-timing in Neuronal Groups. STDP coordinates the dynamics between only neural pairs. Far from being a disadvantage, its action is interesting since it permits to produce a flexible system organization based on many very small *scripts*. Rich information, for instance, can be represented in the network spatially and temporally at the lowest level by neural pairs built into hierarchies of assembled complex patterns. Following this idea, STDP has for some respect similar attributes with the Bayesian rule. The timing-dependent synaptic activation of the neuron $neuron_{post}$ by the activating neuron $neuron_{pre}$ can be devised as a conditional rule between the two units in the form of a script, for instance: if $neuron_{pre}$ fires at time t_{pre}, then $neuron_{post}$ fires at time $t_{post} = t_{pre} + \Delta t$.

Such pragmatic timing rules between two neurons represent the smallest "quanta" of information possible to encode. They form, inside the network, a repertoire of primitives that can be used for example to model the motoric system in order to constitute a "grammar" of action primitives (Rizzolatti & Arbib, 1998). It follows that more complex rules – or abstract representations of actions and behaviors – can be constructed from the dynamical assembling of these basic pairs into long-range spatio-temporal clusters of very short conditional codes see Fig. 2.

Moreover, if the network is sufficiently large, neurons may exhibit non-trivial connection assembling apparent to a spontaneous activity or to self-organization. This spontaneous activity may enable then the system to process information beyond its availability as exposed in Fig. 2: the activation at precise timing of particular neurons before t_1 generates the reconstruction process of the *whole* spatio-temporal cluster till t_5. The sequence is retrieved from partial information and one may not see the complete sequence if the first neurons do not fire. To some extent, this retrieval of spatio-temporal patterns can be seen as a "trajectory attractor". Once an event is re-activated, it follows the ongoing synchronization of other units firing—the idea behind chaos itinerancy (Kuniyoshi, Yorozu, Inaba, & Inoue, 2003; Tsuda, 1991; Tsuda, Fujii, Tadokoro, Yasuoka, & Yamaguti, 2004). To pursue our analogy with Bayesian statistics, we can interpret the long-range spatio-temporal clusters as *enfolded* causal chains of scripts (as for Markovian tree) e.g.,

if X fires at t_1, then Y, Z fire at resp. t_2 and t_3, and if Y, Z fire at resp. t_2 and t_3 then etc . . .

In this fashion, the neuronal groups form hierarchies of different level descriptions set up from their basic neural bricks in a bottom–up fashion, in line with recent biological data supporting that the motoric system is organized into hierarchical representations (Lestou et al., 2008). Although some computational frameworks have been proposed to model hierarchical representations for action representation (Demiris & Simmons, 2006; Wolpert, Doya, & Kawato, 2003; Wolpert, Ghahramani, & Flanagan, 2001), they do not emphasize the importance of timing as the neuro-science dynamical systems viewpoints do (Edelman, 1987; Kelso, 1995; Rabinovich, Varona, Selverston, & Abarbanel, 2006; Tsuda, 1991), which we think important for its functioning. Besides it, polychronization of neural pairs might establish a "vertical association" between parallel neural processes to represent actions and to re-enact them.

2.4. Retina model

The vision map is coarsely inspired by the serial processing done in the retina transforming a video sequence into spike trains (Wohrer, Kornprobst, & Viéville, 2006). The successive transformations done in the ganglion layers realize a complex filtering on the images into a set of spikes (see Fig. 3). The mechanism discriminates both the spatial and the temporal information from a scene, apparent to a spatio-temporal convolution filtering: the combination of a contrast enhancement on the images (a spatial edge refining) sensitive to "temporal edges" (a processing apparent to optical flow making the neurons trigger to moving objects only). The filtering is modeled with the two-pass Gaussian blurring in the spatial domain in the two layers V_A and V_B (detecting the smooth areas in the image) convolved temporally (detecting temporal edges, high-pass temporal filter):

$$V_{A_{\mu_A,\sigma_A}}(x, y, t) = \frac{1}{2\pi\sigma_A^2}\, e^{-f(x,y,\mu_A)/2\sigma_A^2}.$$

$$V_{B_{\mu_B,\sigma_B}}(x, y, t) = \frac{V_{A_{\mu_A,\sigma_A}}}{2\pi\sigma_B^2}\, e^{-f(x,y,\mu_B)/2\sigma_B^2}.$$
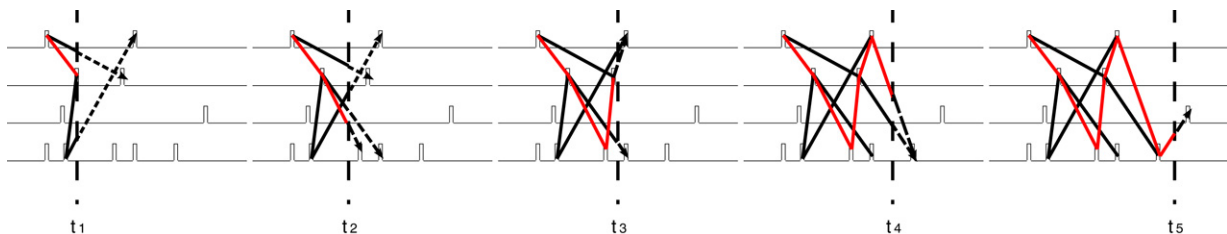
(5)

Fig. 2. Self-organization and hierarchical representations. High level neuronal groups are formed from the assembling of lowest level neural pairs self-arranged in a bottom–up fashion. Accurate information about the timing and the location of spikes permit to have a flexible system organization to represent complex sensori-motor structures into long-range spatio-temporal clusters. One of the ideas behind polychronization of Izhikevich (2006).
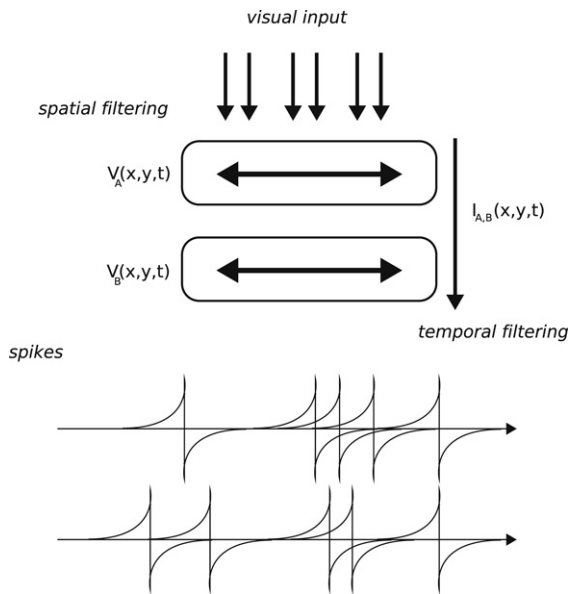
Fig. 3. Retina model. Rough architecture of the information treatment done in the retina transforming a video sequence into spikes trains. The two layers V_A and V_B process a successive spatial filtering of the visual inputs which are then convolved together producing a spatio-temporal filtering of the visual inputs into spike trains.

with $f(x, y, z) = (x-z)^2 + (y-z)^2$, μ and σ respectively the center and the variance of the convolution with values $\{\mu_A, \sigma_A\} = \{0, 1.0\}$ and $\{\mu_B, \sigma_B\} = \{0, 0.5\}$. The pixel output of the retina layer located at $\{x, y\}$ provides the excitatory current distribution, $I_i(t)$ to its associated neuron i Eq. (1) with:

$$I_i(t) = V_{B_{\sigma_B, \tau_B}}(x, y, t) - V_{A_{\sigma_A, \tau_A}}(x, y, t+1). \tag{6}$$

3. Experiments of eye–hand coordination and grasping

We reproduce the experimental series conducted by Rizzolatti et al. (1996) illustrating the qualitative aspects of mirror neurons and of canonical neurons: inter-modal binding, action representation and action understanding with temporal constraint. These neurons combine visuo-motor properties to represent one action sequence and to fire at precise timing. In our experiments, we investigate the conditions for such situation to arise in a network of spiking neurons that would lead from the temporal linkage between the visuo-tactile maps to actions representation. We count, to this end, on the regulating roles of STDP and of the body (embodiment) to coordinate the neuron dynamics to the timing integration among the maps.

In the first part, we conduct some repeated experiences of visually perceived acts (i.e., seeing and touching one object) to be mapped in the neural system in the form of linked visuo-tactile representations (encoding both vision and tactile information). Over time, we expect the network to acquire the direct matching from behaviors to neural dynamics. As the representation of physical actions is fetched into the network as multi-modal patterns, it would be possible then to access one modality from the activation of the other. In the second part, we consider how the network integration will permit to access one missing modality from the activation of another one for instance to the understanding of actions performed by others, when no tactile information is received.

3.1. Description of the experiment

The experiment consists of repeated executions of the action sequence associated to grasping (i.e., reaching the cup – time to contact – grasp – tearing) till convergence of the network dynamics to a stable organization. A schematic of the experiment is presented in Fig. 5. During physical interactions, the vision and tactile maps receive their respective information; a time-line of the action sequence "seen" from the sensor maps is presented in Fig. 4. The vision map receives the pre-processed signals from the retina, the tactile map receives the associated force gradient at the object surface. Since we consider the timing information particularly important (e.g., time-to-contact), we assume that this information can either come from the fingertips or from the object surface. We chose the latter solution for practical reasons.

The visual map is composed of 5400 neurons receiving the output signal from the pixel associated with; 90×60 camera resolution which corresponds to 5400 pixels. Their value, binarized after being filtered by the retina layers, are then normalized to [0; 20] and fed to their relative neurons input I in Eq. (1). The correspondence equation between the pixel coordinates $\{i, j\}$ to the neural index, *neuron_ID*, is: *neuron_ID* $= i \times 90 + j$.

Information at the tactile sensor surface is sampled with 0.5 mm^2 resolution into a data grid of 1000 samples (details in A.1). Each sample is associated to one neuron in the tactile map and 1000 neurons are composing the tactile map. Below 1 N force pressure, the tactile neurons receive no input value from their corresponding sample, $I = 0$, whereas above 1 N force pressure, each sample triggers their corresponding neuron with $I = 20$ (see Eq. (1)). The two maps are composed of eighty percent of the neurons present in the whole network, all excitatory (6400 units). The other twenty percent (2000 units) are inhibitory neurons added to stabilize the global system. Finally, each neuron, either excitatory or inhibitory, is initially connected to one hundred others arbitrarily selected within the global network with equal synaptic weight ($c_{init} = 5$). Under this condition, the ensemble forms a sparse network with no functional connections before learning. Since the inhibitory neuron activity do not correspond to any representational patterns, we will not display them in the following sections.

3.2. Learning eye–hand visuo-tactile coordination

Experiencing grasping. Perceiving the inter-modal and temporal correlations is an important factor for learning the significance of actions, i.e., to decompose the sequential order of goal-directed movements and to distinguish between them the means and ends (Falck-Ytter, Gredeback, & von Hofsten, 2006). Implicit temporal relations between neural dynamics permit to recognize and to detect if one particular event in the sequence order has occurred or not. This role is held in our experiments by STDP to the detection of coincidental events at the neural level.

For instance, when experiencing grasping, the neurons trigger to the synchronous spatio-temporal events of the scene extracted from the sensors. The firing patterns permit to deduce the signals' "hidden causal structure" that discriminates the whole sequence order into action primitives from its preparation to its execution.

The processing done in the visual map and in the tactile sensors permits easily to distinguish the three phases of grasping and to discern its temporal structure (see Fig. 4):

phase 1 *reaching*, the period before contact with the object,
phase 2 *grasping*, the time to contact and touching,
phase 3 *tearing*, the period when the person handles the object.

The events in the two maps can serve then to construct time-based conditional rules through STDP. The saliency map produced from the retina (Section 2.4), for instance, plays a major role to filter all the static objects from the scene: the output of the retina gets the vision neurons to be sensitive to the moving objects only. Hence, before the grasping at $t = t_0$, the device is completely filtered from the scene and only the hand motion (i.e., its spatial contour) is actively retrieved at $t = t_1$. This interval corresponds

temporal structure of grasping

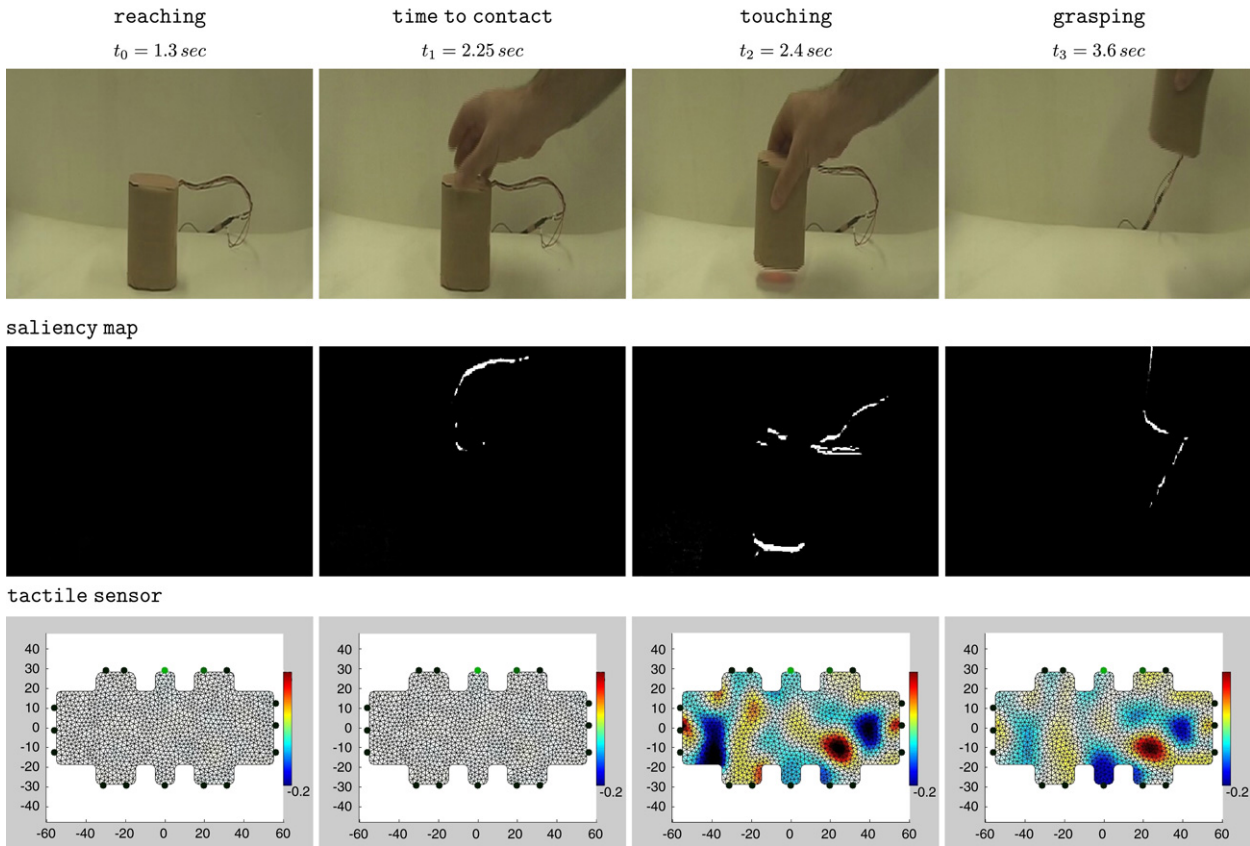| reaching	time to contact	touching	grasping
$t_0 = 1.3\,sec$	$t_1 = 2.25\,sec$	$t_2 = 2.4\,sec$	$t_3 = 3.6\,sec$

saliency map

tactile sensor

Fig. 4. Temporal structure of grasping. Grasping sequence seen from the retina and of the skin sensors. Before grasping ($t < t_1$), the retina detects only the temporal changes about the hand motion in the direction of the cup: the spatial information about the cup is filtered. When grasping the object ($t = t_1$), joint detection of hand motion contingent to the cup motion and the tactile activity corresponding to a coordination in the neural dynamics (synchronization among the maps). Temporal rules about the sequential order of the event are then associated to a neural representation into the network.
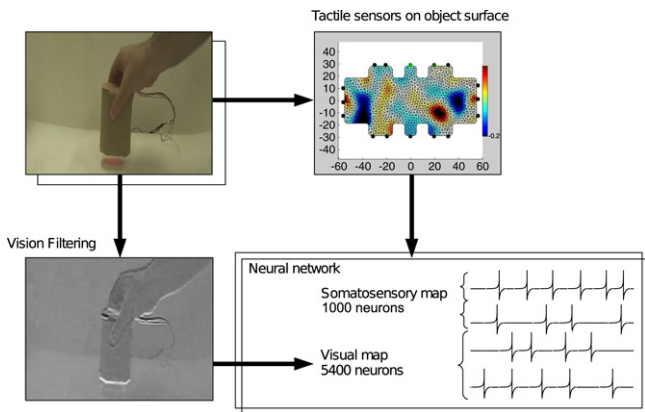

Fig. 5. Schematic of the experiment. The experiencing of co-occurrent visuo-tactile perception during grasping (in the upper-left corner) by the network (bottom-right corner) is done by receiving the incoming information from the camera and from the pressure sensitive device.

to the first phase of the action sequence. The following one stands for the period of time-to-contact at $t = t_2$ and of grasping, when the hand induces some involuntary small perturbances and position changes in the object. These disturbances, accurately detected in the two maps, constitute a unique event distinguished in both maps as the inter-modal grouping of the "hand-device" representation. We analyze in the next section how the network structures its dynamics to categorize the information coming from the sensory inputs.

Network structuring. When experiencing grasping, the network structures its dynamics following the STDP rule: the neuron firing within $\Delta t = 40$ ms latency are wiring together. The temporal coherency between the neurons is the principal factor for their linkage; the location of the neurons is taken into account through the synaptic conduction latency between two spikes. Following this, the neurons can have therefore either short- and long-range connections which can then support segregation within the maps and integration between them. Over time, they may form coherent pairs and clusters associated to the particular experience of grasping with other neurons belonging to the same map or to others. It is this aspect of the network, its functional integration, that we want to analyze in Fig. 6: Fig. 6(a) displays the evolution of the synaptic weights distribution during the learning stage and Fig. 6(b) reproduces the distribution of the paired neurons belonging to the same map or to different maps. This second measure tries to capture quantitatively the network's structural organization evolution. On the one hand, the network level of intra-modal specialization, I_{intra}, corresponds to the information processed between the neurons of the same maps (i.e., the number of synaptic links). On the other hand, the network level of inter-modal integration, I_{inter}, corresponds to the information exchanged between the neurons belonging to different maps (i.e., the number of synaptic links). The third graph in Fig. 6(c) plots the connection matrix between the pre- and post-synaptic neurons belonging to the two maps.

An interpretation of the graphs can be given as follows. The distribution of the neurons' synaptic weights, sharply centered around the unique value 6.0 in Fig. 6(a) in red, indicates a
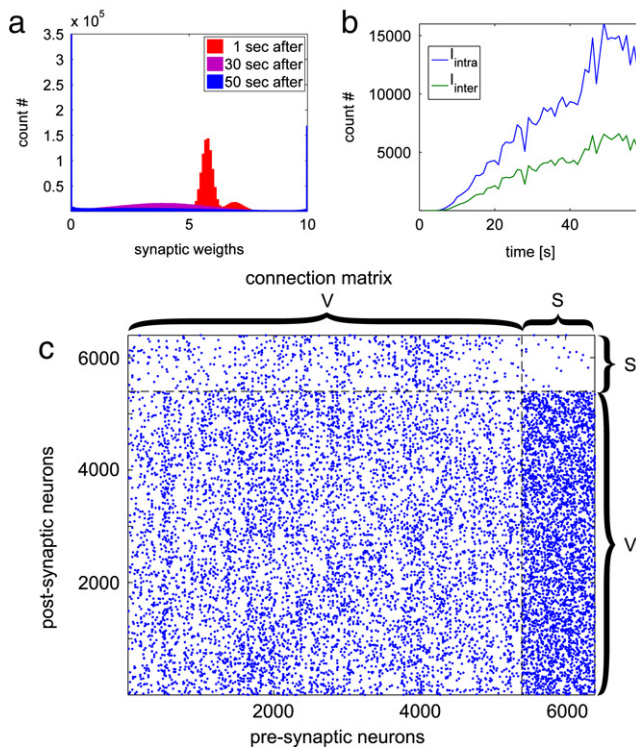
Fig. 6. Network structural organization evolution. Histogram of the synaptic weights (a). Evolution distribution rates of maps' interaction among neurons belonging to the same map I_{intra} and interaction of neurons belonging to other maps I_{inter} (b). In (c), the connection matrix between the pre-synaptic and the post-synaptic neurons belonging to the somatosensory map (S) and to the visual map (V). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Spike rate and neural dynamics in the visual and tactile maps during physical interactions (resp. top and bottom). The whole action sequence is exposed from seeing, reaching the cup, the time to contact, then touching and grasping (see also Fig. 8). In red (resp. cyan), the synaptic activation from the neurons of the vision map (resp. tactile map). Processes done in the network are constituted from the parallel neural firings between the two maps. The visuo-haptic patterns are not randomly activated but synchronized and functionally assembled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

uniform structure: the neural network, before enaction, has no particular organization. The neurons – which are at first arbitrarily connected – will bring forth then an intricated structure in the network similar to a spontaneous random-like activity (Poisson-like distribution). This situation, however, dramatically changes at the beginning of the learning stage when the network receives the contingent visual and tactile information. In this stage, a hierarchical structuring of the network is operating dynamically inside and between the maps, see Fig. 6(b). The vision and tactile maps are self-organizing their internal dynamics while during the same time, at the network level, the maps are also exchanging information.

This stage corresponds to a complexification of the network exhibiting a high degree of specialization *inside* the maps (i.e., $I_{intra} \gg 0$) and a high degree of integration *between* the maps (i.e., $I_{inter} \gg 0$). This situation is shown by the different evolution of the two curves of I_{intra} and I_{inter} engaged inside the network in (b): two processes are taking place, one "horizontal", inside the maps and the other "vertical", between them. Furthermore, these processes are also asymmetric since the density distribution of the synaptic connections from the neurons of the somatosensory map to the neurons in the visual map, $I_{S \rightarrow V}$, is slightly denser than in the opposite direction, $I_{V \rightarrow S}$ (c). A structuring of the network is at work, not symmetric, that corresponds to a developmental stage of functional integration.

Enacted experience. After the learning stage, we reconduct the whole action sequence from reaching to grasping (see Fig. 4) and analyze the network response. The neural activity in the vision and tactile maps during the enacted experience are plotted in Fig. 7 where the black line at $t = 2.3$ s. indicates the precise time-to-contact. The red lines define the synaptic links starting from the pre-synaptic neurons of the vision map, whereas the lines in cyan
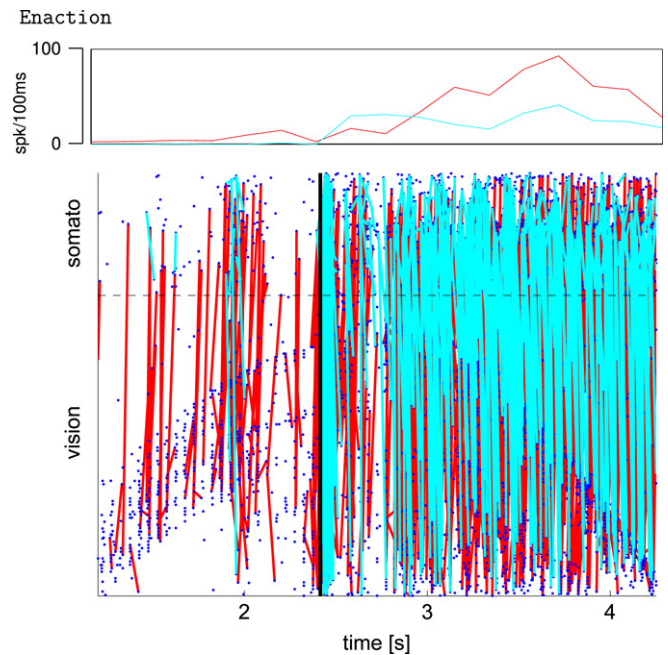
represent the synaptic links starting from the neurons of the tactile map. The spike rate of the correspondent map's distribution with the same color code is plotted in the upper-side region with a resolution of 100 spikes per 100 ms.

Despite the very broad activity in the network, one can observe a clear discrimination of the neural activity before and after grasping. The network has learnt to categorize the saliency from the visual and the tactile inputs and to distinguish the events between grasping and not-grasping. It has therefore detected the proprioceptive-visual contingency of self-produced actions. A result that concords with the analysis done in the previous paragraph [cf. Fig. 6(a)–(c)] showing that the synaptic pairing is conducted between neurons of the two maps with long-range and inter-modal synaptic coupling. In this stage, the two maps intensively exchange information. For instance, some neural spikes in the visual map triggered by external stimuli produces the indirect firing of neurons in the tactile map not necessarily associated to an input stimuli.

In that respect, we can say that the system produces some neural activity not associated to any stimulus. Some activity in the maps is therefore virtual or *simulated*; e.g., see the erroneous triggering of some neurons from the tactile map by some neurons from the vision map *before* any contact for $t < t_1$ in Fig. 7, it does not correspond to any real haptic information but to fake ones (virtual haptic perception). Hence, the network combines stimuli-based dynamics with self-activated cross-modal dynamics. In that sense, the system is not passively receiving information from the environment but is *active* by going beyond the available information. An important feature of the brain dynamics in general and of the MNS in particular which is to simulate fictious perceptions from partial stimuli (Barsalou, 2008; Kent & Lamberts, 2008; Ramachandran & Blakeslee, 1998).

Simulation and anticipation. We analyze now the neural activity in the two maps for the narrower interval at the time-to-contact
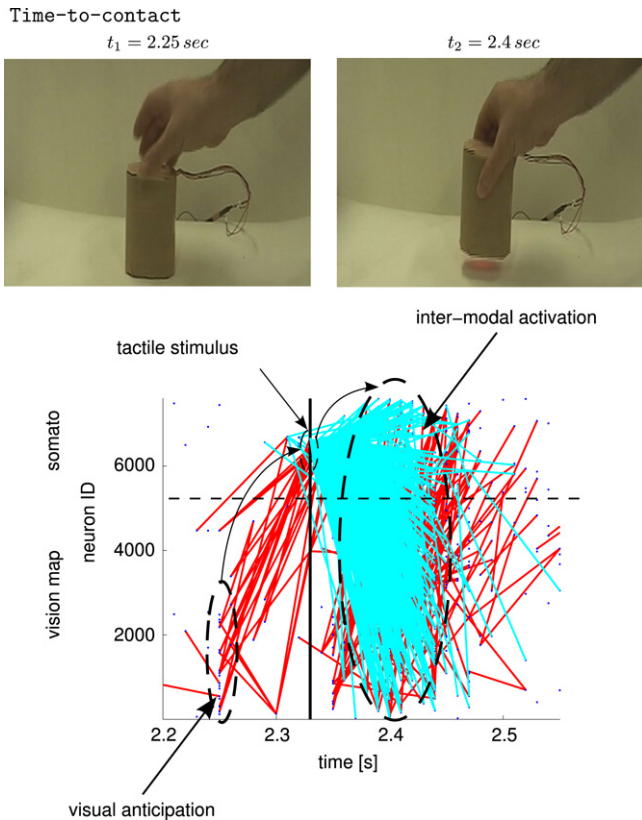
Fig. 8. Neural dynamics of the visuo-tactile maps during physical interactions. In red (resp. in cyan) the synaptic activation from the neurons of the vision map (resp. the tactile map). At time-to-contact ($t = t_1$), the retina *anticipates* only the temporal changes about the hand motion in the direction of the cup: the spatial information about the cup is filtered. When grasping the object ($t = t_2$), joint detection of hand motion contingent to the cup motion and the haptic activity corresponding to a coordination in the neural dynamics (synchronization among the maps). Temporal rules about the sequential order of the event are then associated to a neural representation into the network. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$t \in [t_1; t_2]$ in Fig. 8. Interestingly, two clear intervals separate the neural activity before and after contact (as previously observed) where the first interval is bounded between t_1 to the time-to-contact and the second one from time-to-contact to t_2. In the first interval, one can observe that the specific neural pattern in the visual map, associated to the hand-reaching-the-device (dashed ellipse at $t = t_1$), is triggering the neurons of the tactile map *before* its effective activation from the tactile stimuli at $t = 2.32$ s. (vertical bar). Hence, the vision map anticipates the next state of the tactile map with tens of milliseconds in advance (70 ms). The network is therefore not simply receiving the co-occurrent stimuli but is also coordinating: the visual map predicts the expected activity in the tactile map before its eventual realization; a similar behavior with the one of spatial-visual proximity, observed and stated by Berthoz (1997), which is a form of *anticipated contact* with the area of the body that will be touched (Rizzolatti & Sinigaglia, 2006). Then, if the prospected event is effectively realized – which means that we grasp the device – we will have the tactile neurons conform to the expectation. The result is the visuo-tactile pattern-to-pattern linkage and rewarding through STDP by the increasing of their synaptic weights.

The second interval starts from the time-to-contact (represented by the vertical bar) and ends at t_2 (dashed ellipse) when we effectively grasp the object. In this stage, the tactile stimuli by the fingers' pressure on the device trigger the correspondent neurons of the tactile map. This particular tactile pattern is associated to a
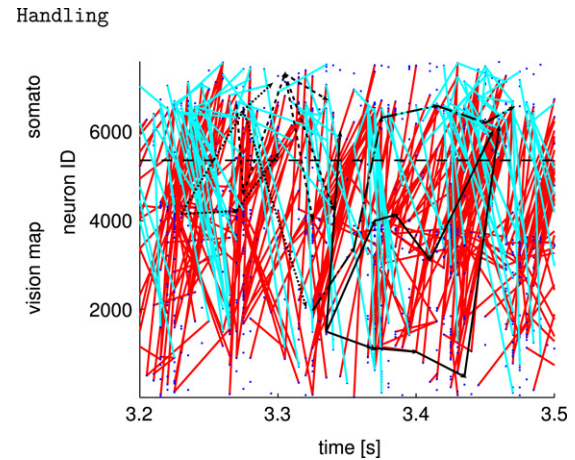

Fig. 9. Causal chain extracted from the visuo-tactile maps. The chains are assembled from the neurons of the two maps. Representation is global and corresponds to the functional integration of all the modalities.

large neural band over the whole network at $t = t_2$. Again, activation is cross-modal between the two maps but this time in the opposite direction starting from the tactile map. Hence, the network can be assimilated to a time-based associative memory, bidirectional, linking the contingent neural activity between the two maps into one event in the neural space.

We can view therefore the network as a corpus of coincident detectors – in the place of neural scripts – always re-enacting and simulating the expected contingent stimulus from the environment. In an information theoretical viewpoint, these features are particularly important for the computational capabilities of this class of network. From a developmental viewpoint, they might underly the ground for contingency detection between different modalities which is argued to pave the way to self-awareness, self-perception and social cognition (Hiraki, 2006; Nadel et al., 2005; Prince & Hollich, 2005; Rochat & Striano, 2000; Watson, 1994).

Retrieval as re-activation. We exposed in the previous section how the two maps exchange information with precise temporal relations, however we might wonder if they form coordinated and coherent activity (i.e., long-range functional integration) during the grasping and handling periods. To answer this question, we propose to analyze the neural dynamics corresponding to these periods during contact for $t > t_2$. We plot them for the interval $t \in [3.2; 3.5]$ in Fig. 9 with the same color code used in Figs. 7 and 8 for the synaptic links.

During handling, the visual map and the tactile map are interacting with each other dynamically and coherently: the parallel processes done between the two maps, which means the inter-modal neural pairing, are also assembled into consistent long-range clusters dynamically constructed, concordant with the perceptual stimuli. The processes produce the particular perception–action matching corresponding to handling-an-object: they get intertwined to each other so that they form effective causal chains and contingency detectors between stimuli from the same modality or to different ones (see the plain and dashed lines in Fig. 9). These links make the two maps to "mirror" their activities: enaction re-activates the familiar tactile patterns which, at their turn, fire back to the visual map. Reversely, neurons from the visual map trigger some tactile neurons which activate others in one or the other map. Thus, perceptual experience in the neural system is not just a simple recall of previously encoded patterns but a constructive process that combines actively information retrieved from the two maps and information received from external dynamics. The neural system re-activates and anticipates neural pairing from different modalities and it is their combination
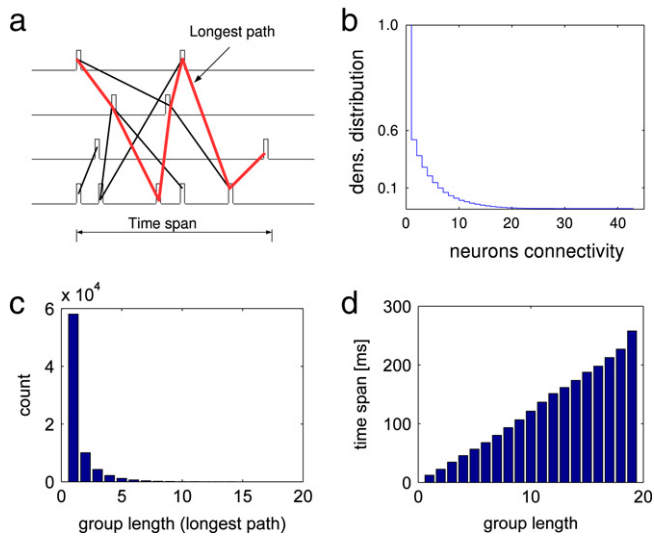
Fig. 10. Cluster statistics. (a) Density distribution of the neuron connectivity (b), ordered by the length of the clusters (c) and by their time span (d) [resp. the longest path of cluster defined and their time span in (a)]. The density of the neuron connectivity follows the characteristic power-law curve typical of small-world networks. The network produces scale-free dynamics. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
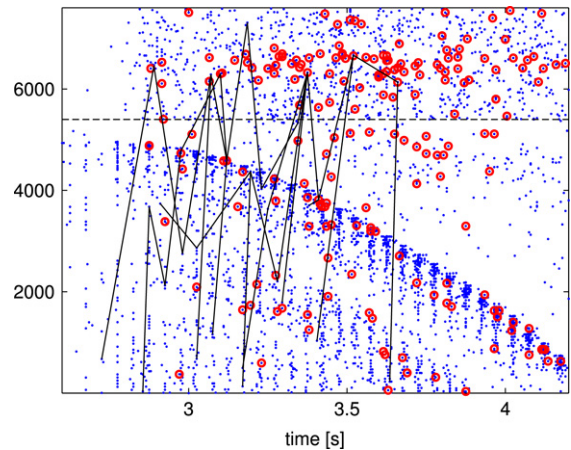

Fig. 11. Critical neurons. We circled in red the neurons with more than ten synaptic connections during the period of tearing [taken from Fig. 10(b)]. They are critical for the functional integrity of the network on which the clusters rely on. We plot also some clusters passing by some of these neurons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that produces the current action representation state inside the network.

Cluster statistics. We study in this section the network statistics. To this aim, we extract from its connectivity matrix the clusters consistent with Izhikevich's definition of polychronized groups: time-locked but not synchronous firing patterns with millisecond order (Izhikevich, 2006). Fig. 10 summarizes the cluster analysis done where one cluster is defined as the neurons grouping with a time span and a path length measured from its *longest* neural path. It corresponds for the cluster in Fig. 10(a) to the neural path colored in red. Fig. 10(b) plots the density distribution of the neuron connectivity inside the network which means the proportion of neurons connected to $1, 2, \ldots, N$ neurons whereas Fig. 10(c) displays the cluster group length distribution and Fig. 10(d) shows the relation between the cluster group length and their time length. Taken together, these measures describe the network functional properties.

For instance, a significant feature is the power-law curves displayed in (b) and (c) – resp. the cluster group length and the neuron connectivity – which reveals the scale-free nature of the dynamics inside the network and informs about its small-world organization. The power laws mean that events in the network are not independent from each other, and a few large events or "hub connectors" (i.e., the tail of the power-law curve) dictate the activity (Buzsaki, 2006). In our experiment, it means that the network produces different description levels of the action across multiple time scales – spanned over 10 to 250 ms, Fig. 10(d) – and assembled dynamically into short- and long-range clusters. Furthermore, their occurrence is not equiprobable; local synchrony (small clusters) more easily arise than global synchrony (long-range clusters) which is more rare thus critical. Complex networks that develop small-world connections are found to generate efficient interregional communication, enhanced signal propagation speed, computational power, and synchronizability (Watts & Strogatz, 1998). It is noteworthy that the MNS has been also found to be organized hierarchically (Lestou et al., 2008) and that *F*5 mirror neurons have been also subdivided into an asymmetric distribution between "strictly congruent" neurons, representing

one third of the population, and "broadly congruent" neurons, representing the two thirds (Gallese et al., 1996).

Similarly, certain neurons are found critical within the network due to their small number in respect to their large number of connections. For instance, less than 5 percent of the neurons possess only more than ten synaptic connections which corresponds roughly to 300 neurons; one third possess more than 5 connections. They represent the "strictly congruent" mirror neurons of the network. We circled them in red in Fig. 11 and drew in black some clusters linking them. As one can see, these neurons follow the trends of the visuo-tactile patterns. They represent therefore the primitives that direct the network's neural activity.

3.3. Understanding actions of others

Reentry, the inter-modal binding of the neural dynamics, let us envision that the neural system can, from experiencing one modality, reconstruct some others unaccessible. For instance, when observing some actions performed by others; the vision information is enough to sense and simulate the missing proprioceptive information. Reciprocally, the manipulation of one object in hidden conditions (e.g., when we close the eyes) permits us to reconstruct its visual primitives (we mentally simulate the missing visual information).

This ability to perceive the action as a whole even from partial information is an important feature of the MNS, it relates to action understanding and to the so-called correspondence problem formulated by Brass and Heyes (2005):

"When observing another person moving, we do not see the muscles underlying their movement but rather the external consequence of that action. So how does the observer's motor system knows which muscle activation will lead to the observed movement? Which relation brings out the correspondence problem between movement observation and motor representation?"

In our experiments, we argue that STDP links the contingent tactile and visual patterns into a unified representation of the action. To demonstrate our hypothesis, we propose to reproduce Rizzolatti's experiment exhibiting the MNS property to trigger with precise timing to observed actions (Rizzolatti et al., 1996). We consider to this end to reconduct the grasping experience in the same experimental conditions as previously done (i.e., the same device to be grasped, the same position on the visual field with the same camera angle as in Fig. 4 with an overall similar execution speed), but this time without providing any tactile information to
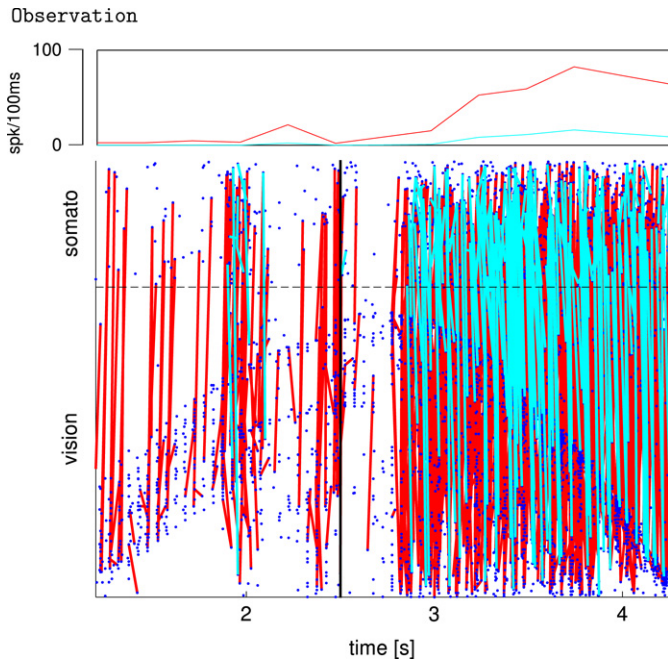
Observation

Fig. 12. Spiking rate and activity of the neural maps in response to action of grasping observed (no tactile information provided) – resp. top and bottom. The neuron-to-neuron synaptic links in red (resp. cyan) correspond to neural activation having for pre-synaptic neurons those belonging to the vision map (resp. tactile map). The black line delineates the time-to-contact. Before contact, the neural activity presents similar activity than during enaction; see Fig. 7. At the time-to-contact and during handling, the visual map activates the dynamics of the tactile map even without any input stimulus. The network simulates and reconstructs the missing modality with precise timing. Such virtual perception in the network reproduces one of the qualitative property of the mirror neuron system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the network. We display in Fig. 12 with the same color code as used in the previous sections the neural map activity in response to the

observed action of grasping, the time-to-contact is at $t = 2.5$ s (plain line).

Before the time-to-contact, during the observation of the hand-reaching-the-object, the neural dynamics exhibit similar patterns as during enaction (see Fig. 7). One might consider that the situation will differ from true enaction at the time-to-contact and during grasping since the network will not receive any tactile information. However, the neural maps do reproduce similar dynamics after $t > 2.8$ s despite the missing modality (to compare with Fig. 7). The visual patterns corresponding to the-hand-grasping-the-object-and-handling-it, activate (in red) the neurons belonging to the tactile map which, as for the case of true enaction, fire back (in cyan) the vision neurons, as if the tactile information were effectively provided. The haptic perception is therefore fictious: because of the massively parallel and bidirectional connections between the two maps, the two modalities are intertwined and accessible to each other. Thanks to reentry, signals that come to one spot from a map return then back to its own map (Edelman, 1987; Edelman & Tononi, 2000). As a result, the network reproduces similar qualitative features of the mirror neuron system.

We plot in Fig. 13(a)–(b) the details of the neural dynamics for the respective period of time-to-contact, $t = 2.5$ s, and during handling, $t > 2.8$ s. At the time-to-contact in Fig. 13(a), even without experiencing any tactile feedback, the network is nevertheless capable to anticipate the precise timing of the missing modality and to reconstruct its dynamics. The simulated modality comes from the contingent cross-modal circuits that were activated during the learning stage and that are reactivated during the retrieval stage. But in contrast to the former situation, the grasping event does not yield to any tactile stimuli. It follows that the inter-modal binding associated to the tactile activation in Fig. 8 is now missing. Despite this disrupting in the information exchange, the perceptual modality is shortly reconstructed during the handling in Fig. 13(b). The recall, based on massive parallel small scripts, is therefore robust. Links from the tactile map to the visual map in cyan show that perception is an active process inside the system whether physical (stimuli-based) or simulated
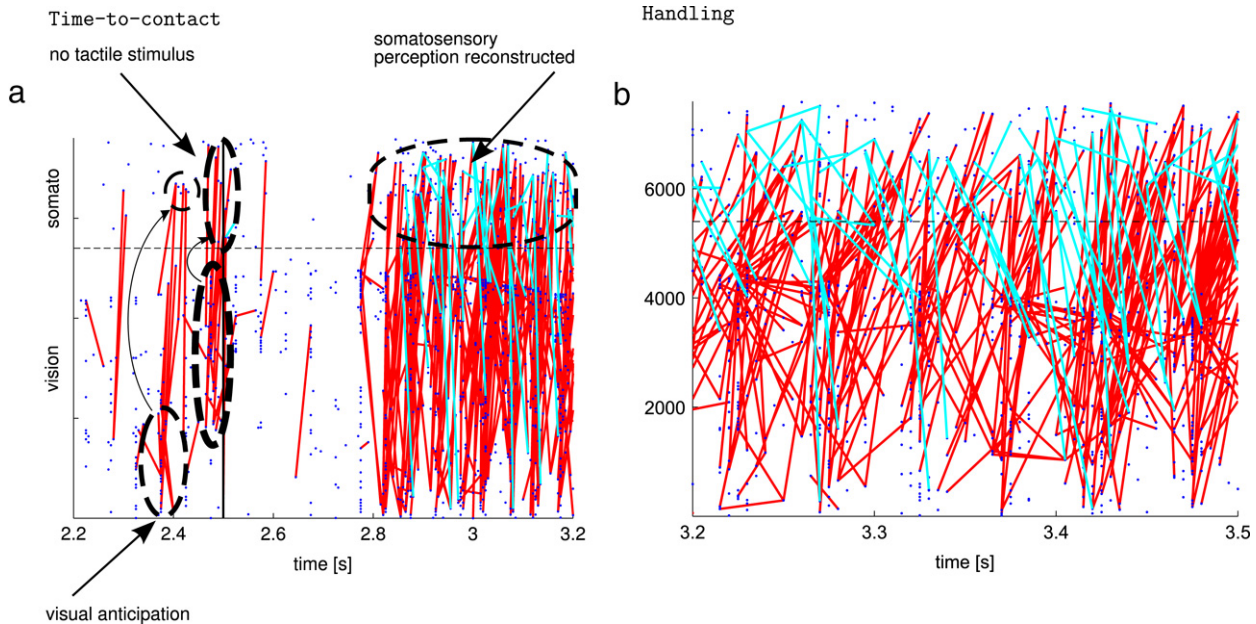

Fig. 13. Reconstruction of the tactile modality from the observed action of grasping at the time-to-contact (a) and during handling (b) details of Fig. 12 with the same color code. The network anticipates the tactile stimulation at the precise time to contact in (a) however, the expectation is not fulfilled which differs from the case of effective enaction when expectation of the tactile stimulation is rewarded (see Fig. 8). Soon after during the handling of the device for $t > 2.8$ in (a) and (b), the visual patterns permit to simulate and to reconstruct the missing tactile perception. Links from the tactile to the visual map in cyan show that perception is an active process inside the system whether physical (stimuli-based) or virtual (not stimuli-based) emerging from the mirroring between the two maps.

(not stimuli-based) emerging from the mirroring between the two maps and using the same neural pathways.

4. Discussion

We present a biologically plausible mechanism to the representation of one action in a multi-modal neural network based on the learning mechanism of STDP. Temporal structure of complex actions (e.g., grasping) are decomposed with millisecond order precision into ordered sequence of neural rules. The assembling of these very many small scripts from contingent visuo-tactile inputs produce at the body level coherent clusters of hundreds of milliseconds order range expanded in the whole network. These massively parallel connections are reentrant and bidirectional between the two maps. They sustain the information exchange between them and their functional integration. The result is that new cognitive skills emerge at the body scale such as multi-modal integration, anticipation, and simulation of one missing modality (i.e., perceiving virtually its stimulation not from a physical stimulus). The qualitative features of the mirror neuron system.

Neurons that fire contingently, wire together. In a developmental stage, the co-occurring information received from vision and tactile input during the grasping sequence permit to shape the neural system structural organization. Enaction produces the epigenetic control of the network anatomy into an efficient organization. To this end, we emphasize the role of the body to structure the neural dynamics and the coordination among the maps. Emergent properties such as reentry, scale-free dynamics and computational capabilities appear *thanks to* the neural network embodiment. Selection from situated action is decided through reentry relying on the pairing of critical neurons or "hub connectors" which are relatively few inside the network but possess a high number of connections. They represent the "strictly congruent" mirror neurons in the *F*5 area whereas the peripheral neurons represent the "broadly congruent" ones (Gallese et al., 1996).

As a result, the neural system acquires the appropriate perception–action associations reproducing some of the qualitative properties of the MNS: "reverberation" between modalities, firing to executed and to observed actions with precise timing. We speculate that the mirror neuron system is modeled with precise intermodal coupling mediated by the regulatory mechanism of STDP for action representation, anticipation and action understanding. Our hypothesis is plausible since STDP governs the neural dynamics in various brain areas including the neo-cortex and the *F*5 area but restricted since it does not explain how abstract action goals might be interpreted. Lestou et al. (2008) and Rizzolatti and Craighero (2004) distinguish nevertheless these two separated functions in some recent studies.

The findings of Lestou et al. (2008) suggest that processing in the ventral premotor cortex (*PMv*) may mediate the exact copying of complex movements (e.g., precise grasps, finger prehension and whole hand prehension), whereas processing in the parietal and superior temporal areas may support the interpretation of abstract action goals (e.g., reach, take, hold, and tear). Our model might provide to this end a plausible mechanism of the *PMv* area based on dynamical system viewpoints (Edelman & Tononi, 2000; Izhikevich, 2006; Kuniyoshi et al., 2003; Tani, Ito, & Sugita, 2004; Tsuda et al., 2004) with respect to computational approaches (Kawato, 1999; Miall, 2003; Oztop et al., 2006; Wolpert et al., 2003).

To conclude, in a developmental viewpoint, the spike-timing-dependent plasticity might provide the sufficient neural basis for babies to sense contingency, what Watson (1994) hypothesized to be the premises for body representation, self-perception and the discrimination between self and others which are the conditions for developing social capabilities (Hiraki, 2006; Nadel et al., 2005; Rochat, 1998, 2003). In this line, we hypothetize that the present neural architecture might support some principles for cognitive development and social competences necessary for communication by means of gesture and language (Falck-Ytter et al., 2006).

Acknowledgment

Appendix

A.1. Haptic sensor

Our original tactile sensor consists of a pressure sensitive conductive rubber sheet (sensor sheet), with 16 electrodes placed only on its boundary. Electrical current is injected into the sensor sheet in 16 different patterns using 16 different pairs of neighboring electrodes. Each pattern of current injection produces a unique potential distribution based on the internal resistance of the sensor sheet. These potential distributions are sampled on the boundary of the sensor sheet using the rest of the electrodes, to produce a total of 208 potential measurements for all the current injection patterns i.e. a data frame.

The sampled data frame is used to estimate the sensor's internal resistance distribution based on an inverse analysis method called Electrical Impedance Tomography (EIT). Based on this concept, any change in the resistance distribution of the material through pressure, stretch or other tactile stimuli can be detected throughout the sensor sheet. Since there are no wirings in the internal part, a thin, deformable and deformation sensitive tactile distribution sensor is formed [cf. (Alirezaei, Nagakubo, & Kuniyoshi, 2007a, 2007b)]. Applications of the tactile distribution sensor include easy implementation over complex 3D surfaces, implementation over stretching areas such as robot joints, and detection of sophisticated tactile stimuli involving skin deformation.

The sensor has a reasonable sensitivity threshold and can detect forces bigger than 1 N. Also, it can detect tactile stimuli which stimulate more than 1% of the sensor area which is an acceptable resolution for our experiment. An FEM model with 2261 elements, each having an associated resistance value, is used to estimate and visualize the resistance distribution of the sensor [Fig. 4]. The estimated resistance values of the 2261 elements are used as the tactile input and directly fed into the neural network. These resistance estimations are produced at a rate of around 20 frames per second. We then down-sampled the resolution by two to one thousand elements to be sent to the units of the tactile map.

References

Abbott, L., & Nelson, S. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, *3*, 1178–1182.

Alirezaei, H., Nagakubo, A., & Kuniyoshi, Y. (2007a). A deformable and deformation sensitive tactile distribution sensor. In *Proc. of 2007 IEEE intl. conf. on robotics and biomimetics* (pp. 1301–1308).

Alirezaei, H., Nagakubo, A., & Kuniyoshi, Y. (2007b). A highly stretchable tactile distribution sensor for smooth surfaced humanoids. In *IEEE-RAS 7th intl. conf. on humanoid robots*.

Aoki, T., & Aoyagi, T. (2007). Synchrony-induced switching behavior of spike pattern attractors created by spike-timing-dependent plasticity. *Neural Computation*, *19*, 2720–2738.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Berthoz, A. (1997). *Le Sens du Mouvement*. Odile Jacob, english translation Harvard Press.

Bi, G., & Poo, M. (1998). Activity-induced synaptic modifications in hippocampal culture, dependence of spike timing, synaptic strength and cell type. *Journal of Neuroscience*, *18*, 10464–10472.

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

Buzsaki, G. (2006). *Rhythms of the brain*. Oxford University Press.

Changeux, J., & DeBevoise, M. (2004). *The physiology of truth : Neuroscience and human knowledge*. Cambridge: Belknap Press of Harvard University Press.

Corbetta, D., Thelen, E., & Johnson, K. (2000). Motor constraints on the development of perception–action matching in infant reaching. *Infant Behavior and Development*, *23*, 351–374.

Demiris, Y., & Simmons, G. (2006). Perceiving the unusual: Temporal properties of the hierarchical motor representations for action perception. *Neural Networks*, *19*, 272–284.

Edelman, G. M. (1987). *Neural darwinism: The theory of neuronal group selection*. New York: Basic Books.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness (Consciousness)*. Basic Books.

Falck-Ytter, T., Gredeback, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neurosciences*, *9*(7), 878–879.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*, 23–48.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 593–609.

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

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

Hiraki, K. (2006). Detecting contingency: A key to understanding development of self and social cognition. *Japanese Psychological Research*, *48*(3), 204–212.

Izhikevich, E. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, *14*, 1569–1572.

Izhikevich, E. (2006). Polychronization: Computation with spikes. *Neural Computation*, *18*, 245–282.

Izhikevich, E. M., Gally, A. J., & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex*, *14*, 933–944.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *89*, 460–471.

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.

Kent, C., & Lamberts, K. (2008). The encoding-retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*(3), 92–98.

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

Lestou, V., Pollick, F., & Kourtzi, Z. (2008). Neural substrates for action understanding at different description levels in the human brain. *Journal of Cognitive Neuroscience*, *20*(2), 324–341.

Meltzoff, A., & Moore, K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75–78.

Miall, C. (2003). Connecting mirror neurons and forward models. *NeurReport*, *14*, 2135–2137.

Nadel, J., Prepin, K., & Okanda, M. (2005). Experiencing contingency and agency: First step towards self-understanding in making a mind?. *Interaction Studies*, *6*(3), 447–462.

Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, *2126*, 1–18.

Prince, C., & Hollich, G. (2005). Synching models with infants: A perceptual-level model of infant synchrony detection. *Journal of Cognitive Systems Research*, *6*(3), 205–228.

Rabinovich, M., Varona, P., Selverston, A., & Abarbanel, H. (2006). Dynamical principles in neuroscience. *Reviews of Modern Physics*, *78*, 4.

Ramachandran, V., & Blakeslee, S. (1998). *Phantoms in the brain*. HarperCollins.

Rizzolatti, G., & Arbib, A. (1998). Language within our grasp. *Trends in Neuroscience*, *21*, 188–194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

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

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

Rizzolatti, G., & Sinigaglia, C. (2006). *Mirrors in the brain*. Raffaello Cortina Editore, english translation Harvard Press.

Rochat, P. (1998). Self-perception and action in infancy. *Experimental Brain Research*, *123*, 102–109.

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, *12*, 717–731.

Rochat, P., & Striano, T. (2000). Perceived self in infancy. *Infant Behavior and Development*, *23*, 513–530.

Song, S., Miller, K., & Abbott, L. (2000). Competitive hebbian learning and through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3*, 919–926.

Tani, J., Ito, M. Y., & Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot experiments using rnnpb. *Neural Networks*, *17*, 1273–1289.

Tsuda, I. (1991). Chaotic itinerancy as a dynamical basis of hermeneutics in brain and mind. *World Futures*, *31*, 105–122.

Tsuda, I., Fujii, H., Tadokoro, S., Yasuoka, T., & Yamaguti, Y. (2004). Chaotic itinerancy as a mechanism of irregular changes between synchronization and desynchronization in a neural network. *Journal of Integer Neuroscience*, *3*, 159–182.

Watson, J. (1994). Detection of self: The perfect algorithm. In S. Parker, R. Mitchell, & M. Boccia (Eds.), *Self-awareness in animals and humans: Developmental perspectives*. Cambridge University Press.

Watts, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, *393*, 440–442.

Wohrer, A., Kornprobst, P., & Viéville, T. (2006). From light to spikes: A large-scale retina simulator. In *Proceedings of int. conf. of neural networks* (pp. 4562–4570).

Wolpert, D., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society*, *358*, 593–602.

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. (2001). Perspectives and problems in motor learning. *Trends in Cognitive sciences*, *5*(11), 487–494.

Zukow-Goldring, P. (2005). Assisted imitation: Affordances, effectivities, and the mirror system in early language development. *From action to language via the mirror neuron system*.