Advanced Robotics, Vol. 20, No. 10, pp. 1105–1126 (2006)

Also available online - www.brill.nl/ar

Full paper

Adaptive body schema for robotic tool-use

COTA NABESHIMA *, YASUO KUNIYOSHI and MAX LUNGARELLA

Laboratory for Intelligent Systems and Informatics, Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan

Received 1 November 2005; accepted 27 March 2006

Abstract—The development and expression of many higher-level cognitive functions, such as imitation, spatial perception and tool-use, relies on a multi-modal representation of the body known as the body schema. Although many studies support the hypothesis that the body schema is adaptive and alterable throughout ontogenetic development, the mechanisms underlying its plasticity have yet to be clarified. Here, we argue that the temporal integration of multisensory information is a plausible candidate mechanism to explain how manipulated objects (e.g., tools) can become incorporated into the body schema. To demonstrate the validity of our idea, we introduce a model of body schema adaptation instantiated in a small-sized, table-top, tool-using humanoid. The robot's task is to learn to reach for and touch a visually salient distant object, first with its 'bare' hand and then — using the acquired know-how — with a reach-extending tool (a stick). Our experimental results show that in order to successfully causally relate and integrate vision, touch and proprioception, and to learn to use the tool, timing is of crucial relevance. On a more general note, this study also suggests that synthetic modeling might not only be a valid avenue towards getting a better grasp on results provided by neuropsychology and neurophysiology, but also a powerful approach for building advanced tool-using robots.

Keywords: Body image; tool-use; adaptation; visuo-tactile integration; developmental robotics.

1. INTRODUCTION

In humans and other primates, many fundamental abilities depend on the availability of an internal (e.g., neural) representation of the body's current posture and spatial extension. Humans, for instance, imitate other people by projecting visually observed behaviors onto their own motor capabilities and action possibilities. This non-conscious mapping between visual perception and action, which is also known as 'transferability' [1], relies on the integration of multi-modal sensory information (e.g., visual, tactile and proprioceptive, i.e., information originating

*To whom correspondence should be addressed. E-mail: nabesima@isi.imi.i.u-tokyo.ac.jp

from the body itself) into a common body-centered representation. Interestingly, such representation is not limited to the individual body parts, but can be shown to include the body's peripersonal space, i.e., the space immediately surrounding the body [2], pointing to its crucial role for spatial perception. This is plausible if one considers that in order to guide movements through space the brain must constantly monitor limb positions and make predictions about their future positions in relation to surrounding objects that can be reached (most of the involved bodily operations remain outside conscious attention).

Another important instance of how the body representation dynamically extends and incorporates pieces of the environment, that 'normally' would not be considered part of it, is tool-use (e.g., Ref. [3]). Tool-use is the ability possessed by humans and other animals (e.g., birds, horses and chimpanzees) to use different tools to manipulate objects and, hence, 'move beyond' the limits set on their action space by the length of their limbs or the type of their end-effectors. Neurophysiological evidence indicates that after repeatedly using a tool, changes in the body representation (of humans and trained monkeys) occur and over time the tool is perceived as if being an integral part of the body. The phenomenon describing the extension of the body representation to include non-corporeal objects in the peripersonal space, such as tools or prosthetic devices, has been extensively reported in the literature [4–7] and lies also at the center of this article.

Two distinct and complementary definitions of 'body representation' exist (see Refs [8] and [9] for critical reviews): (i) the body schema, i.e., a non-conscious neural map of the spatial relations among the parts of the body which integrates multi-modal sensory information (e.g., visual, somatosensory, tactile and proprioceptive) [10–12] and (ii) the body image, a consciously manipulable version of the body schema [13] which relates to the phenomenal experience of one's own body (self-awareness). The latter representation of the body refers 'to a conscious visual representation of the way the body appears 'from the outside' according to Haggard and Wolpert [36], and is also concerned with a more conceptual understanding of the body in general [14]. It is still unclear how to bridge the explanatory gap between these two concepts, but some attempts are currently under way (e.g., Ref. [9]). In this paper, we tackle the problem by looking at it through the lens of developmental robotics [15]. Our methodology is double-pronged and synthetic: we first extract design principles or mechanisms from neurophysiological and neuropsychological findings, and then instantiate those principles in a behaving artificial system. Although the focus of this article is on the body schema, no apparent conceptual obstacles exist to integrating body image in our framework. In other words, we suggest that synthetic modeling might provide a viable solution to the problem of creating a coherent and consistent theory to connect the two notions of body representation.

In what follows, we first briefly describe the notion of body schema by drawing upon findings from neurophysiology, psychology and robotics. Our core assumption is that the body schema plays a pivotal role for robotic systems intended to interact with human beings or to display high-level cognitive functionalities. From

the perspective of our synthetic approach, the study of tool-use by means of robots has two distinct advantages: (i) tool-use extends the robot's manipulatory abilities and, thus, improves its interaction with the surrounding world (e.g., humans) and (ii) the effects of tool-use adaptation on any kind of internal representation are easily accessible, and have a clearly visible outcome which can be detected, studied and hopefully understood. In Section 3, we outline a novel biologically inspired model of body schema adaptation, which we instantiate in a real robot that has to learn to use a simple tool in order to extend its reach. We proceed by reporting our results and discussing them in the light of how our model might be helpful to understand how tools are actually incorporated into the body representation of an actor which manipulates and interacts with its local environment. This study describes a synthetic model for physiology and presents a novel approach towards tool-use in robots and, thus, might be of value for physiologists, psychologists and roboticists.

2. PERSPECTIVES ON THE BODY SCHEMA

In this section, we present a view on the body schema which hinges on studies from the areas of neurophysiology and neuropsychology, and briefly survey some previous research applying the concept of body schema to robots.

2.1. Phantom limbs

The term body schema was initially introduced by Head and Holmes [10] to describe (i) the mapping of proprioception and efference copies (that is, copies of motor commands) onto body posture and movements, and (ii) the mapping from tactile sensation to its originating location on the body surface. We call this original definition the 'geometric body schema'. The concept of body schema — and the one of body image, for that matter — has been substantially enriched by the discovery of the phantom limb phenomenon [16]. This phenomenon can be observed in amputees who have lost their limbs either due to accident or to disease and who find themselves feeling somatic sensations (such as pain, tickle or itch) originating from their amputated (i.e., non-existent) limbs (hence the name phantom limbs) [17]. Although it would help clarify the debate concerning the question of whether the body schema is innate or acquired, no conclusive evidence exists showing that phantom limbs also occur in the case of congenitally missing limbs — the literature on the issue is ambiguous and controversial [9].

2.2. Plasticity

Recent studies shed new light on the plasticity of the body schema. Phantom limbs, for instance, can be treated with mirrors which give amputees relief from the disturbing presence of the lost limb [18]. This finding demonstrates that the body schema is alterable even in adults, and treated patients can learn to 'rationally'

accept the conflict between visual and 'phantom' somatosensory information and to adjust to the new body. Based on the results of these studies, it is probably correct to assume that the body schema is plastic and alterable, constantly updated by sensory feedback (e.g., Ref. [7]). More directly relevant to our study is a recent report on the plasticity of a class of bimodal neurons in the monkey's intraparietal cortex which typically respond to somatosensory and visual stimulation from the monkey's hand. Surprisingly, after the monkey is trained to retrieve distant objects using a rake, the visual receptive fields of the bimodal neurons expand to include the entire length of the tool used and the mere sight of the tip of the rake is sufficient to activate the bimodal neurons [5]. This result not only confirms the hypothesis that the body schema is adaptive, but it also shows that it can actually extend to incorporate tools and other inanimate objects.

Some other remarkable instances of body schema and body image extension are described in the literature. A representative example is shown in Fig. 1 (see Ref. [19]). In this particular experiment, a subject sits at a table while hiding both hands. The experimenter simultaneously taps and strikes one of the subject's hands and an unreachable point on the table. Astonishingly, after a minute or so, the subject starts feeling as if the stimuli emanates from the table — despite the table being an inanimate object! This verbally reported extension of the body to include the table is a striking example of (conscious) 'body image extension' (we are not aware of any work addressing this issue, but we hypothesize that it would probably require more time for the body schema to extend given the typical time constants of neural systems). The study provides yet more evidence that body representations such as the body image or the body schema are plastic and easily alterable — even in adults. Note that despite the proven plasticity of both representations, their underlying mechanisms also exhibit stability in the sense that brain regions seem to be permanently committed to consciously represent body parts [4].

The most plausible implication from all these studies appears to be that the body schema is open to modifications and alterations throughout one's lifetime (modifications that can occur on small time scales). Indeed, during ontogenetic development we experience a continuous and incessant deluge of multisensory

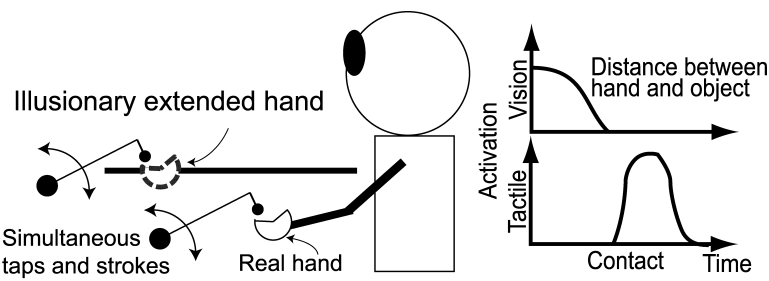

Figure 1. Schematic illustration of the phenomenon of body image extension. If taps and strokes on the table and on the real hand occur simultaneously, then the subject feels the table as being part of her body — she is misled to believe that her hand has actually extended.

information, and the body schema might be constantly affected by the neural integration, re-organization and storage of exteroceptive and proprioceptive sensory feedback.

2.3. Robots

It is interesting to observe that the concept of body schema is closely related to the modeling techniques adopted in traditional robotics where control strategies are often concerned with the establishment of various coordinate systems to represent positions and orientations, and models of forward or inverse mappings among such coordinate systems, e.g., mappings from a body-centered system (e.g., joint coordinates) to a world-centered system [20]. In analogy to the geometric body schema, such conventional kinematic and kinetic models are inalterable (static), and are based on a representation of the robot's body in a Cartesian coordinate system. In robotics, geometric transformations from one space to another are not only used for planning trajectories, but also for integrating different kinds of sensory information, e.g., visual and proprioceptive [21], visual and auditory [22] or visuomotor information [23]. Typically, spatial information is extracted from camera images, or from the intensity or the phase difference between auditory signals, and then mapped onto a Cartesian coordinate frame attached to the body of the robot. Similarly, it is possible to convert proprioceptive information. The Cartesian coordinate system provides a common and uniform ground on which to combine the sampled sensory information.

In contrast to traditional approaches and more in tune with current adaptive techniques to robot control [24, 25] or motor control [26, 27], recent research has attempted to endow robots with some kind of adaptive body schema. Yoshikawa *et al*. [28], for instance, proposed an adaptive body representation system which starting from uninterpreted (raw) visual data first learns to identify spatial regions occupied by the robot's body parts and then builds up a map of the body. Stoytchev [29] proposed a model of body schema extension which, based on the premise that bodily changes can be detected, can autonomously devise a strategy to adapt the robot's kinematic controller to such changes. Moreover, given that the robot 'knows' which object is a tool (tools can be conceptualized as bodily changes), the same author also introduced an approach to represent and learn tool affordances grounded in the robot's behavioral repertoire [30]. Another related piece of work is that by Metta and Fitzpatrick [31] who programmed a robot to poke objects with its arm (without using a tool) and through estimations of the induced optic-flow to discover properties of objects such as rollability.

Our approach differs from previous modeling efforts. The work by Yoshikawa *et al*., for instance, addresses the question of how to acquire a body representation and not how such a representation can be adjusted to incorporate external objects. The work by Stoytchev discusses kinematic changes of the body with regard to tool-use, but not how such bodily alterations can be detected. Kinematic changes of the body involve spatial transformations and, therefore, represent a problem amenable

to mathematical solution. We conjecture that the real problem might actually be how to detect the alterations based on the sensory data. Metta and Fitzpatrick, finally, focused on the extraction of information structure from multi-modal sensation, but neglected adaptation and tool-use. The work presented in this paper shows a novel different stance on tool-use. We suggest that tool-use depends on the coherent unification of spatial and temporal aspects, particularly temporal integration of multisensory information. Our model integrates and complements some of the ideas contained in previous work: kinematic changes are indeed required when learning to use a tool, and information about time is crucial for tool detection, identification and use. We introduce our model in the following section.

3. ADAPTIVE MODEL OF BODY SCHEMA

As already stated before, the body schema is not a static entity, but changes plastically. In this section, we propose a mechanistic model of body schema adaptation which relies on the temporal integration of multisensory information.

3.1. Timing-based model

Our model is based on three main assumptions. The first assumption is that visual and tactile information are integrated in time and space. By 'integration in space' we mean a mapping between the coordinate systems of each sensory modality without the need to postulate a common intermediate frame of reference. This assumption simplifies our model and increases its affinity to temporal integration. Touch occurs when the robot moves its arm so as to approach a target object and eventually contacts the object with its hand. When executing this movement sequence, the robot first obtains visual information while its hand is approaching the target and then tactile information upon contact with the target object. Therefore, visual and tactile sensations are causally related. Moving the hand to approach a target represents the cause and the tactile sensation (contact) represents the corresponding effect. This also means that by associating visual information (available during the approach phase) with tactile information (resulting from the contact with the target object) it is possible to predict the cause of a possible action. Tactile and visual information are also spatially integrated: visual 'contact' (i.e., overlap of hand and target object) and tactile contact co-occur on the same point on the body. Thus, the two modalities are integrated in time and space, and unified as one modality by the event 'contact'.

Our second assumption is that it is possible to store the integrated sensory information in an associative memory so that if one modality is active, the other one is co-activated. In this sense, two spatial locations from which visual and tactile 'contact' information originate are considered to be the same location, only if the available information on the event 'contact' is compatible with the one retrieved from the associative memory. In what follows, such co-occurrence of spatial and

temporal information is called 'rationalization'. The design principle used relates to the concept of causality: the body representation of the actor (trained monkey, human subject or robot) extends only if the temporal–causal relationship among the multiple sensory modalities matches a relationship which the actor has previously experienced and which can be retrieved from memory. Our last assumption is that the robot possesses already an *a priori* kinematic knowledge of its body after learning. This assumption does not exclude ontogenetic development. The focus is on how the robot detects a tool as a modification of its body and how it adapts to such a modification.

3.2. Tool-use model with plastic body schema

We augment our timing-based model of body schema adaptation to include adaptive tool-use. The proposed strategy enables a robot to reach and touch a distant target with a tool by treating the tool as it were the hand. Note that knowledge about the tool is *a priori* unavailable; the robot autonomously incorporates the tool in its body schema and learns to use it.

The proposed strategy is composed of four steps:

(i) Temporal integration. The robot reaches for and touches an object with its 'bare' hand, and learns to temporally associate and integrate visual and tactile information. The learned relationship is stored in an associative memory (Fig. 2).

(ii) After integration. If tactile information is obtained after learning to temporally integrate sensory information, then the robot recalls the associated visual information by retrieving it from memory (Fig. 3).

(iii) Rationalization. If the recalled visual information is consistent with the actually obtained visual information, then the location of visual 'contact' is considered as the location on the hand from which the tactile sensation originated.

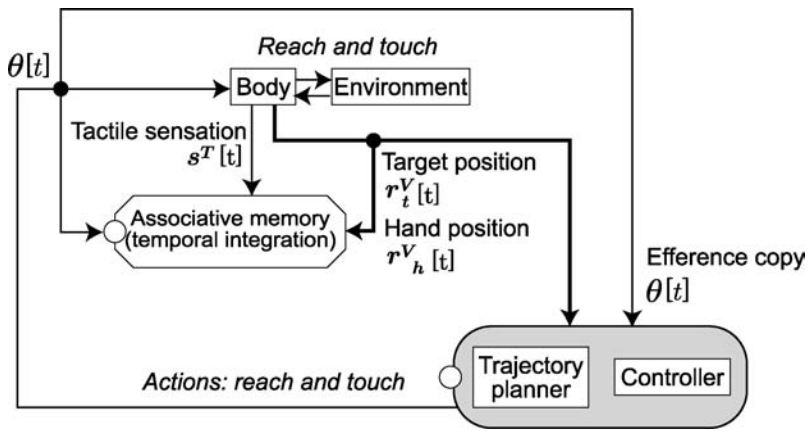

Figure 2. Information flow during learning. Temporal integration of visual and tactile information.

(iv) Learning of controller. If visual 'contact' occurs not on the robot's hand, but on the tool, then the robot is not able to adequately use the tool with the current hand trajectory/position controller. This induces the system to learn a new kinematic controller for the tool.

The detailed information flow during the temporal integration phase is depicted in Fig. 2. Henceforth we refer to $\boldsymbol{r}$ as a vector expressed in a Cartesian coordinate frame attached to the body of the robot. The superscript V denotes visually obtained information and the superscript T tactile information. To reach the target with its hand (step (i) of the strategy described above), the robot needs to execute the following steps. First, the robot generates spatial and kinematic trajectories of its hand $\tilde{\boldsymbol{r}}^{\mathrm{V}}[t]$ starting from the hand's current position $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[0]$ to the target location $\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[0]$; the trajectories are then internally simulated and an executable trajectory is selected using the hand kinematic controller; the robot then actually moves the hand along the internally simulated trajectory $\boldsymbol{\theta}[t]$ (converted from $\tilde{\boldsymbol{r}}^{\mathrm{V}}[t]$ by the controller), and eventually reaches and touches the target. The temporal visual information $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t]-\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[t]$ and tactile pattern $\boldsymbol{s}^{\mathrm{T}}[t]$ obtained are fed to the associative memory which acts as a temporal integrator.

Let us assume now that the robot holds and swings a stick. A tactile sensory pattern $\boldsymbol{s}^{\mathrm{T}}[t]$ is induced when the tip of the stick accidentally hits a target (step (ii)). The visual location of the endpoint of the stick at the time of contact is denoted as $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t_{\mathrm{c}}]$. The temporal pattern of tactile sensation $\boldsymbol{s}^{\mathrm{T}}[t]$ is fed to the associative memory, which recalls the previously learned temporal pattern of the pair of visual signals $\tilde{\boldsymbol{r}}_{\mathrm{h}}^{\mathrm{V}}[t]-\tilde{\boldsymbol{r}}_{\mathrm{t}}^{\mathrm{V}}[t]$. The retrieved pattern is then compared with the ones actually obtained for the stick and the hand, $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]-\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[t]$ and $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t]-\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[t]$. The better fitting pattern, $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]-\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[t]$, is chosen and regarded as corresponding to the previous experience of touching the target with the bare hand. This results in recognizing the endpoint of the stick $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t_{\mathrm{c}}]$ as the current source of tactile signal $\boldsymbol{s}^{\mathrm{T}}[t_{\mathrm{c}}]$, which originally

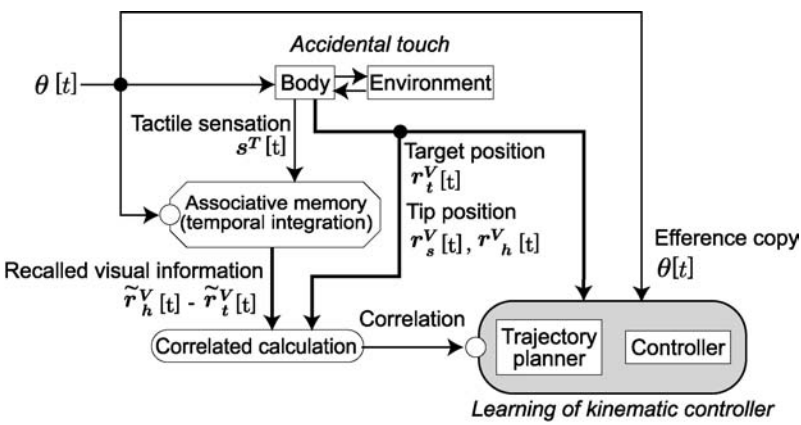

Figure 3. Information flow after learning. Visual information is retrieved from the associative memory and compared with the current visual information.

was $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t_{\mathrm{c}}]$ in the bare hand case. We call this process 'rationalization' (step (iii)). The robot has become 'aware' of the tool, referring to it as if it is the hand.

By adopting this strategy the robot can determine which controller should be employed in order to use the stick (here, the controller of the hand) and to learn the reach-extending controller based on the available visual information $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]$ and the efference copy $\boldsymbol{\theta}[t]$ (step (iv)). After learning, the robot is able to use the tool in the same manner as the hand: the robot first generates a set of paths, internally simulates those paths and finally moves the tool along an executable trajectory. At this point, we presume that the tool has actually become a part of the body and has been integrated in the robot's body schema.

4. IMPLEMENTATION

In this section, we describe our experimental setup and the implementational details of our model.

4.1. Experimental setup

Our experimental setup consists of a small-sized table-top humanoid robot (Figs 4 and 5). The robot is equipped with one color CCD camera, two touch sensors (located in the robot's hand) and a two-link manipulator constrained to move in the horizontal plane. The robot's task is to learn to reach for and touch a colorful object located in front of it, first with its 'bare hand' and then — using the acquired know-how — with a stick. By means of the stick the robot can extend the distal link of its

Figure 4. Robot used in our experiments (left). Snapshots of a selected reaching experiment (right).

manipulator and hence its action space. A contact between the hand and the object or the stick and the object activates the switches in the hand's palm (Fig. 6) whose outputs are fed to the neural system (see below). Note that the activation of the switches is dependent on the target approach direction. The CCD camera is used for the acquisition of the positions of limbs and objects on the plane. We pre-calibrate the distortion of the lens and pre-calculate the transformation from camera-centered coordinates to body-centered coordinates. Both operations are performed using a set of perspective images collected with the camera before performing the actual experiments.

The information entering the system is the following: (i) tactile information $\boldsymbol{s}^{\mathrm{T}}[t]$ originating from the touch sensors, (ii) visual information consisting of the positions of the robot's hand $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t]$, the tip of the stick $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]$ and the target object $\boldsymbol{r}_{\mathrm{t}}^{\mathrm{V}}[t]$, and (iii) proprioceptive information given by the joint angles $\boldsymbol{\theta}[t]$ (the proprioceptive sensation is analogous to efference copies because the robot's actuators are feedforward position-controlled servo motors). For simplicity of implementation, either $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t]$ (bare-hand case) or $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]$ (stick-using case) is input to the system by means of a colored marker detected by image processing. This is justified by the fact that $\boldsymbol{r}_{\mathrm{s}}^{\mathrm{V}}[t]$ is quite distant from $\boldsymbol{r}_{\mathrm{h}}^{\mathrm{V}}[t]$ in our setup and the selection between the two is trivial. The approaching direction $\boldsymbol{\phi}$ of the hand (or the stick) to the target is the angle formed by the line connecting the hand (or the tip of the stick) and a horizontal line parallel to the torso of the robot (see Fig. 5).

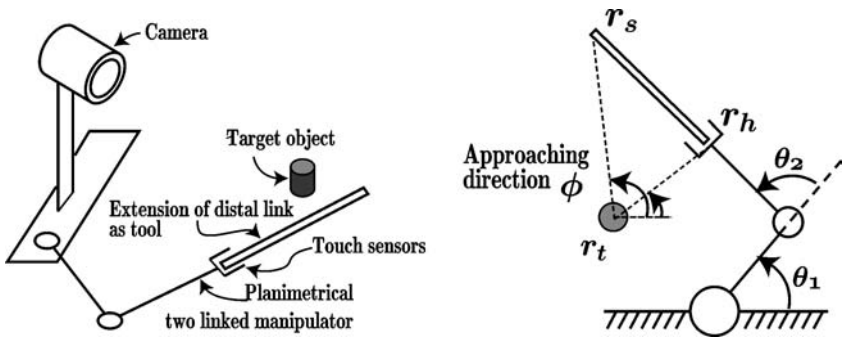

Figure 5. Schematic illustration of experimental setup (left), and variables used in our model (right). See text for details.

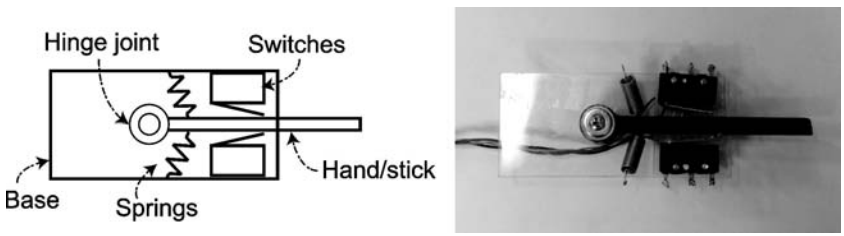

Figure 6. Palm module of the robot. Schematic representation (left) and realization (right).

4.2. Design of spatio-temporal associative memory

The implemented associative memory is shown in Fig. 7. It consists of the combination of two associative memories: a gating neural network (GNN) associating the visually detected target approach direction with tactile information, and (b) a non-monotone neural network (NNN) [32] associating tactile signals with the distance $d[t]$ between hand and target. Both networks are essentially augmented Hopfield neural networks and an incomplete or noisy input can be used to retrieve the stored pattern.

4.2.1. GNN. The GNN is composed of 12 neurons which have the following activation function:

$$y_i = \operatorname{sgn}\left(\sum_j w_{ij} y_j + z_i - \mu\right), \tag{1}$$

where w_{ij} is the connection weight from neuron j to neuron i, y_i is the output of neuron i, z_i the input to neuron i, μ a fixed threshold and $\operatorname{sgn}(x)$ a function returning $+1$ or -1 depending on the sign of the argument x. The GNN obeys an augmented Hebbian-learning rule [33]:

$$w_{ij}^* = w_{ij} + y_i(y_j - y_i w_{ij}), \tag{2}$$

where w_{ij}^* and w_{ij} are the new and the old weight, respectively.

As shown in Fig. 8, the tactile feedback $\boldsymbol{s}^{\mathrm{T}}[t]$ is fed to two sensor neurons. The approaching direction $\phi[t]$ is first weighed by 10 Gaussian kernels fully connected to each other. The weighted values are then fed to 10 hidden neurons (all other neurons in Fig. 8). The purpose of the GNN is to associate tactile sensation and the visually determined approach direction of the hand relative to the target object, implying a static and spatial (but not temporal) association between vision and touch.

4.2.2. Connection between networks. The outputs $y_i[t]$ of the tactile neurons of the GNN (neurons connected to the tactile sensors) are eventually summed

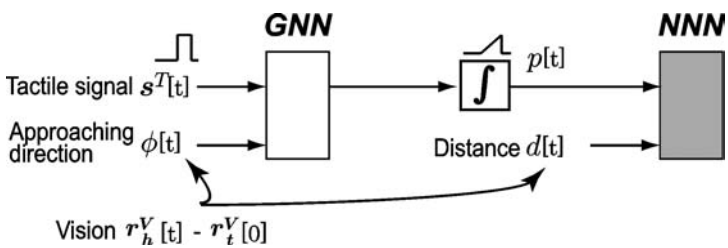

Figure 7. Visuo-tactile spatio-temporal associative memory. The memory is composed of a GNN and a NNN.

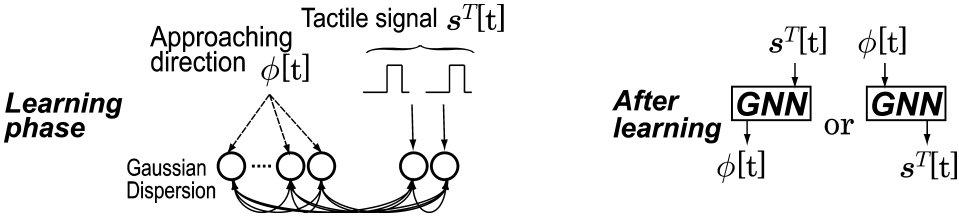

Figure 8. GNN. Information flow during learning (left) and after learning (right). $\phi[t]$ is the approaching direction, $s^{\mathrm{T}}[t]$ is the tactile pattern.

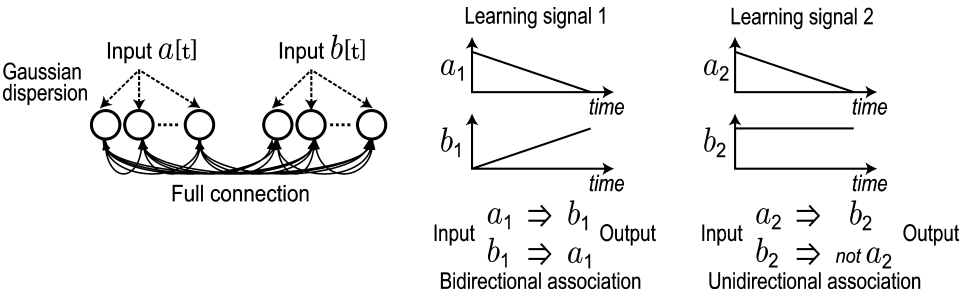

Figure 9. Ideal requirement of a Hopfield-like neural network (left) for association of temporal patterns. Unidirectional association (center). Bidirectional association (right).

according to:

$$p[t] = \sum_{t' < t} \sum_{i \in \text{tactile sensors}} y_i[t'], \tag{3}$$

where $p[t]$ is the sum over time up to time t. Note that, because the output of the tactile sensors is either 0 (absence of collision) or 1 (presence of collision), $p[t]$ represents a signal containing a slope and is extended in time (see Fig. 7). The purpose of this sum is to transform tactile patterns $s^{\mathrm{T}}[t]$ (localized in time) into a signal varying over time, the reason being that it is easier for the NNN to learn temporally extended input patterns. The Hopfield-like network used for the association of visual and tactile patterns is a sequential memory and learns patterns at each time step. Partially correct input patterns are sufficient for the system to retrieve stored patterns. It means that the patterns learned at each time step should (ideally) differ from each other. Patterns containing a slope satisfy this requirement. In this case the network functions as a bidirectional associative memory (e.g., learning pattern 1 in Fig. 9). In contrast, if one of the learning patterns is constant, the network works as a unidirectional memory (e.g., learning pattern 2 in Fig. 9).

4.2.3. NNN. The NNN is composed of 40 neurons. As shown in Fig. 10, 20 neurons of the NNN code for the distance $d[t]$ between hand and target, whereas the other 20 neurons code for the tactile pattern $p[t]$. Both are weighed by 20 Gaussian kernels. The activation dynamics of the NNN is governed by the following

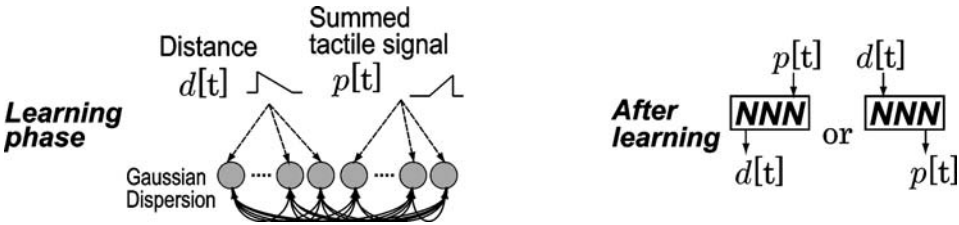

Figure 10. NNN. Information flow during learning (left) and after learning (right).

Table 1.
Parameters of the non-monotone neural network

Δt	τ	τ'	λ	α'	c	c'	h	κ
0.05	2	30 000	0.2	0.2	50	10	0.5	−1.0

set of equations:

$$\tau \frac{\mathrm{d}u_i}{\mathrm{d}t} = -u_i + \sum_j w_{ij} y_j + z_i, \tag{4}$$

$$y_i = f(u_i), \tag{5}$$

$$f(x) = \frac{(1.0 - \mathrm{e}^{-cx})(1.0 + \kappa \mathrm{e}^{c'(|x|-h)})}{(1.0 + \mathrm{e}^{-cx})(1.0 + \mathrm{e}^{c'(|x|-h)})}, \tag{6}$$

where τ is an update time constant, u_i is the internal potential of neuron i, w_{ij} is the connection weight from neuron j to neuron i, y_i is the output of neuron i and $z_i = \lambda \gamma_i$ is an external input to neuron i (as described in Ref. [32]); for all neurons λ is constant throughout the learning phase in order to realize a bidirectional associative memory. The learning rule of the NNN reads:

$$\tau' \frac{\mathrm{d}w_{ij}}{\mathrm{d}t} = -w_{ij} + \alpha \gamma_i y_i, \tag{7}$$

where τ' is the learning time constant, γ_i is the learning signal for neuron i and $\alpha = \alpha' x_i y_i$ is the learning coefficient (α' was fixed). For all values of the parameters, refer to Table 1.

These coupled networks can effectively learn the association among spatio-temporal patterns. Because of the Hebbian-like learning, the tactile and visual patterns are fed to the networks only during the learning phase. Due to their associative capabilities, when one modality is fed to the system, the pattern of the other modality can be recalled.

4.3. Design of kinematics learning

In our experiments the tool is a stick attached to the distal link of the arm of the robot. The Jacobian matrix and the forward kinematic equations are given only for

the hand. The relation of joint angles $\boldsymbol{\theta}$ and the position of the end effector $\boldsymbol{r}$ (hand or tip of tool) can be expressed as:

$$\delta \boldsymbol{r} = \boldsymbol{J}(\boldsymbol{\theta})\delta\boldsymbol{\theta}, \tag{8}$$

$$\boldsymbol{J}(\theta) = \begin{bmatrix} -l_1 s_1 - l_2 s_{12} & -l_2 s_{12} \\ l_1 c_1 + l_2 c_{12} & l_2 c_{12} \end{bmatrix}, \tag{9}$$

where J is the Jacobian, l_i $(i = 1, 2)$ is the length of link i, $s_1 = \sin\theta_1$, $c_1 = \cos\theta_1$, $s_{12} = \sin(\theta_1 + \theta_2)$ and $c_{12} = \cos(\theta_1 + \theta_2)$. The above equation is specialized for the hand and the tip of the stick to yield the following equations:

$$\delta \boldsymbol{r}_{\mathrm{h}} = \boldsymbol{J}_{\mathrm{h}}(\boldsymbol{\theta})\delta\boldsymbol{\theta}, \tag{10}$$

$$\delta \boldsymbol{r}_{\mathrm{s}} = \boldsymbol{J}_{\mathrm{s}}(\boldsymbol{\theta})\delta\boldsymbol{\theta}, \tag{11}$$

where $\boldsymbol{r}_{\mathrm{h}}$ and $\boldsymbol{J}_{\mathrm{h}}$ are the position and the Jacobian matrix of the hand, and $\boldsymbol{r}_{\mathrm{s}}$ and $\boldsymbol{J}_{\mathrm{s}}$ are those of the stick. Note that the only difference between the hand and stick is their length. It seems a reasonable assumption because an object held in the hand is equivalent to an enlarged or deformed hand. The relationship between the Jacobians is straightforward:

$$\boldsymbol{J}_{\mathrm{s}}(\boldsymbol{\theta}) = \boldsymbol{J}_{\mathrm{h}}(\boldsymbol{\theta})\boldsymbol{J}_{\mathrm{c}}. \tag{12}$$

Equation (12) means that a robot can learn to use a stick if it can learn the mapping from one Jacobian to the other. We substitute (12) into (11), rearrange the equation with $\boldsymbol{J}_{\mathrm{h}}(\boldsymbol{\theta})$ and the displacement $\delta\boldsymbol{r}_{\mathrm{s}}$ of the performed movement $\delta\boldsymbol{\theta}[t]$, and obtain:

$$\boldsymbol{\Theta}' = \boldsymbol{J}_{\mathrm{c}}\boldsymbol{\Theta}, \tag{13}$$

$$\boldsymbol{J}_{\mathrm{c}} = \boldsymbol{\Theta}'\boldsymbol{\Theta}^{+}, \tag{14}$$

$$\boldsymbol{\Theta}' = \left[(\boldsymbol{J}_{\mathrm{h}}^{-1}(\boldsymbol{\theta})\delta\boldsymbol{r}_{\mathrm{s}})[t]\ (\boldsymbol{J}_{\mathrm{h}}^{-1}(\boldsymbol{\theta})\delta\boldsymbol{r}_{\mathrm{s}})[t-1] \cdots \right], \tag{15}$$

$$\boldsymbol{\Theta} = \left[\delta\boldsymbol{\theta}[t]\ \delta\boldsymbol{\theta}[t-1] \cdots \right], \tag{16}$$

where '+' denotes the pseudo-inverse.

As it is linear and constant, the matrix $\boldsymbol{J}_{\mathrm{c}}$ should be relatively easy to learn. We thus use a simple two-layered perceptron whose input and output are $\boldsymbol{J}_{\mathrm{h}}^{-1}(\boldsymbol{\theta})\delta\boldsymbol{r}_{\mathrm{s}}[t]$ and $\delta\boldsymbol{\theta}[t]$, respectively. The perceptron works as the extensional transformation $\boldsymbol{J}_{\mathrm{c}}^{-1}$ for the controller of the hand $\boldsymbol{J}_{\mathrm{h}}^{-1}$.

5. EXPERIMENTS

In the first experiment, the robot reaches and touches a target with its 'bare' hand. The robot consequently experiences multi-modal sensory information which is integrated spatially and temporally, and stored in the associative memories. To generate the reaching movements, we use quadratic Bezier curves describing the path connecting the hand and target. The hand reaches the target by approaching it from four different directions ($\phi[t_{\mathrm{c}}] = 60, 90, 240, 270°$; t_{c} is the time of contact).

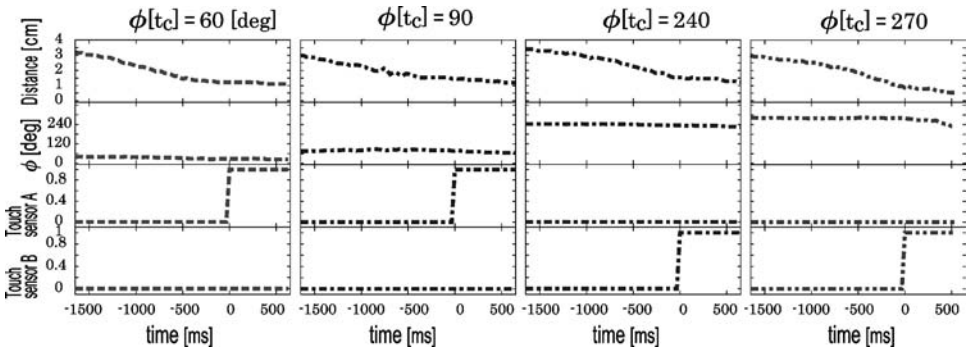

Figure 11. Spatio-temporal patterns generated when reaching target with the 'bare' hand. First row: distance between hand and target object. Second row: approach direction $\phi[t]$. Third and fourth rows: activity of touch sensors A and B. The approaching directions at the time of contact $\phi[t_c]$ are 60, 90, 240 and 270°.

The robot first simulates all movements, then selects a possible trajectory and eventually executes the movement (Fig. 4, right). The expression of this behavior leads to a set of visuo-tactile patterns: the cases of $\phi[t_c] = 60$, 90, 240 and 270° are shown in Fig. 11. Each pattern represents an attempt to approach the target. The top row is the distance $d[t]$ between hand and target. The second row is the approach direction $\phi[t]$. The two bottom rows show the response patterns of the tactile sensors A and B located in the robot's palm. The patterns are shifted in time so that the time of contact is 0 ms. Because the visual patterns are cluttered after contact, they are delayed by 1500 ms and then fed as learning signals to the associative memory. The activation patterns of the NNN are updated until its state stabilizes for every input and for each time step. Similarly, the GNN learns at every time step the spatio-temporal patterns of its input.

In the second experiment, we make the robot hold and swing a stick. When the robot accidentally hits the target with the tip of the stick, it obtains concurrent tactile and visual information. The consequent tactile activation patterns are then fed to the associative memory, and the visual patterns $d[t]$ and $\phi[t]$ are extracted from the NNN through the GNN. If the visual and tactile feedback is temporally consistent with the stored information, the robot becomes 'aware' of the extension of its hand and starts learning a reach-extending controller for the stick. Typically, in less than 60 s the robot learned to use the stick as a tool as if it were its hand. Two input and output patterns are shown in Fig. 12: if putative tactile signals are fed into the associative network, the robot can recall the visual images of approach and contact. Here, the inputs of the GNN are the step waveforms which are activated at time 0. Those inputs are summed by (3) and fed to the NNN as signals with slopes. Two input patterns in Fig. 12 indicate those slopes. One corresponds to the activation of tactile sensor A, whereas the other relates to the activation of tactile sensor B. The recalled patterns were time-correlated with the visual patterns obtained when the robot swung a stick and hit the target. The time-correlations with and without

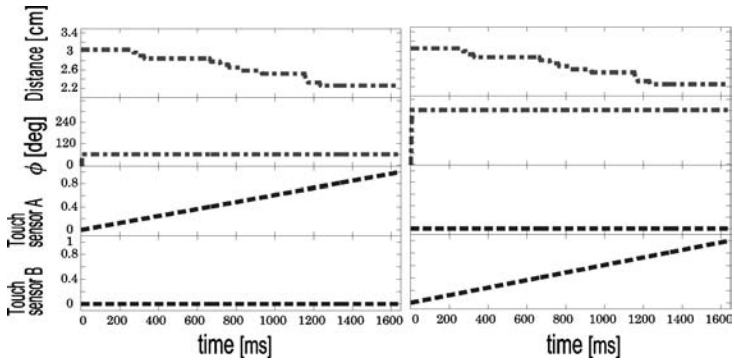

Figure 12. Visual patterns recalled by the associative memory (NNN) when a tactile pattern is used as input. First row: distance between object and tip of tool. Second row: approach direction $\phi[t]$. Third and fourth rows: time-integrated activity of touch sensor A (left) and activity of touch sensor B (right).

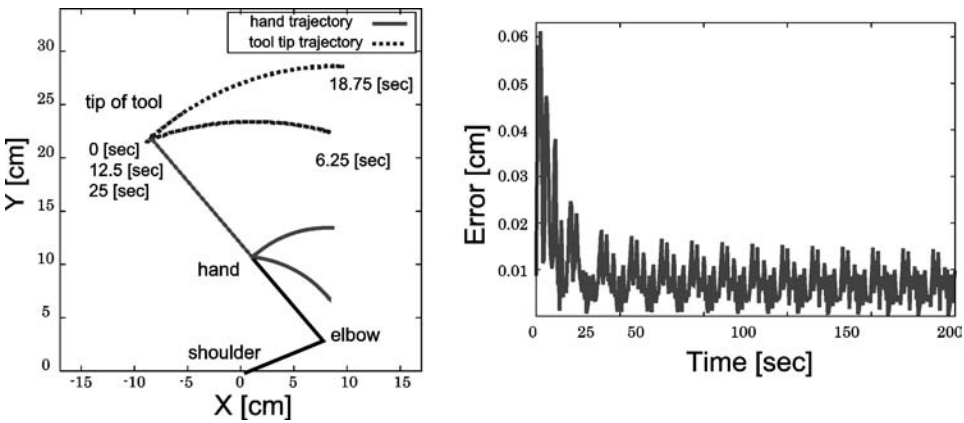

Figure 13. (Left) Trajectories of hand and tool tip during learning inverse kinematic transformation (one learning episode took approximately 25 s). (Right) Learning error (sampling frequency is 2 Hz).

the tool were consistent. It is the temporal consistency and the spatial difference between the hand and the tip of the tool that drive the system to learn the Jacobian.

During the learning of the controller, the robot individually moved its joints for 25 s (as can be seen in Figs 13(left) and 14). The resulting visual patterns $\delta \boldsymbol{r}_{\mathrm{s}}[t]$ as well as the perceived movement $\delta \boldsymbol{\theta}[t]$ were fed to the NNN every 500 ms. The experiment was repeated 50 times with a fixed learning coefficient. The learning converged rapidly as depicted in Fig. 13(right). The ripples in the learning error indicate that once the tip of the tool touched the object, the robot did not stop moving, but kept oscillating to and fro the object. For any given $\boldsymbol{J}_{\mathrm{s}}$, our experimental results $\boldsymbol{J}_{\mathrm{c}}$ are in agreement with our theoretical predictions.

Successful learning means that the robot can use the stick as if it were the hand by merely using the learned Jacobian $\boldsymbol{J}_{\mathrm{c}}$. The same quadratic Bezier curves obtained for the hand are in fact also used to generate reaching movements with the stick. The

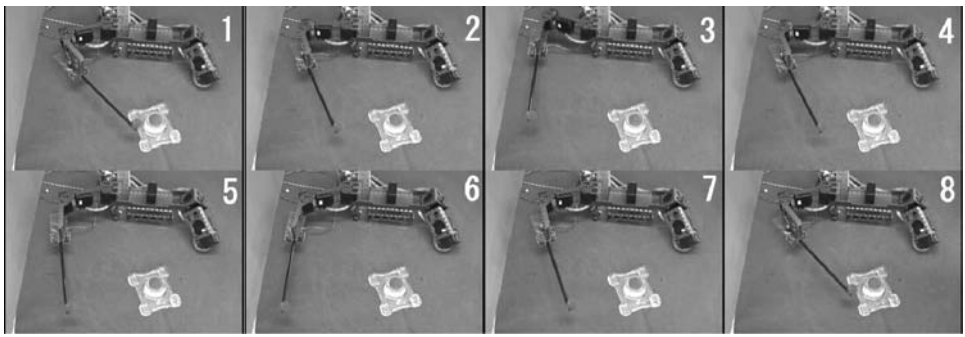

Figure 14. Experiment of inverse kinematics learning.

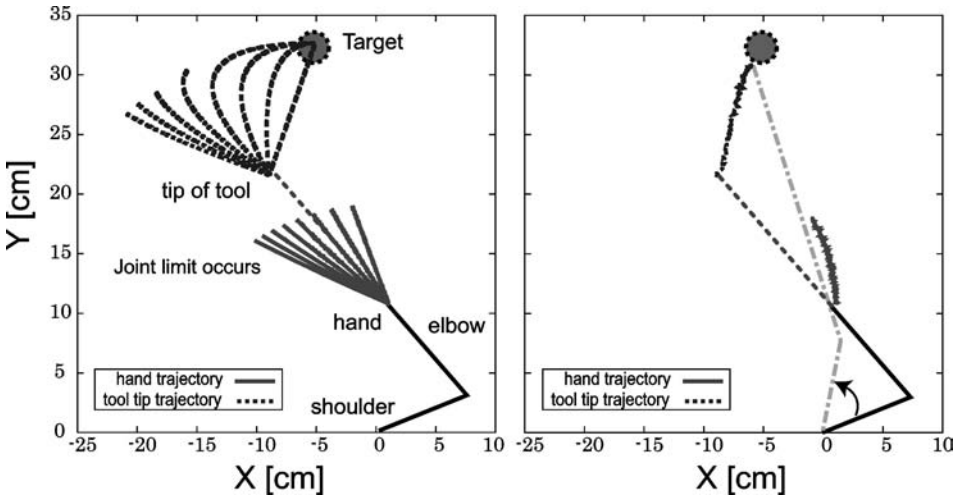

Figure 15. Reaching a target unreachable without a tool. (Left) Simulated trajectories: the lines ending halfway indicate rejection at the ending points due to the joint limit; the lines that reach the target are paths that can be used to perform a movement. (Right) Actually performed movement.

movements are first simulated and then after selecting an appropriate trajectory, they are actually executed. Figure 15(left) reproduces some of the simulated trajectories. The robot was able to judge whether the target was reachable by the hand or if the stick had to be employed. Here, it could successfully opt to use the tool and execute one of the possible paths. The trajectories of the hand and stick tip are depicted in Fig. 15(right).

6. DISCUSSION

In alignment with recent neuroscientific evidence, we suggested that temporal integration of visual and tactile information might be a key mechanism underlying body schema adaptation. By drawing on this mechanism, we introduced a model of body schema adaptation which we instantiated in a small-sized table-top humanoid

robot. The model was tested in an experimental scenario in which the robot had to learn to reach for and touch a visually salient object lying in front of it. To successfully perform in this — admittedly simple — learning task, we endowed the robot with the following 'learning' strategy: it first learned the temporal relationship (i.e., timing) between visual and tactile information, and then based on the acquired 'know-how', if the object was out of reach of the hand, the robot learned to use a stick to reach for and touch the object. A successful performance in this case was only possible if the body schema was extended to include the stick.

The first issue that needs to be discussed concerns the selection of the point(s) of visual attention. In our study, we simplified this problem by attaching colored markers to the hand, the tip of the tool and the target object, making them visually salient, and thus identifiable and segmentable with a simple algorithm. If we consider a more realistic model, we have to take into account automatic attention of those points. We justify our choice by noting that although attention is a relevant and highly investigated problem involving both low-level and high-level cognitive processes (e.g., Ref. [34]), it is not of the essence for the current implementation of our model. There are many instances where visual attention plays a role of course and we will address the issue in the future. Consider, for instance, a situation in which multiple moving objects co-exist or a case in which a 'featureless' or inconspicuous point is crucial for the functionality of a tool (e.g., a small lever hidden below a table). Which point of attention should the system choose? More specifically, should attention be directed to the hand or to the stick, or the object itself? And in which order? One possible solution would be, as attempted here, to visually attend to the object whose motion is synchronized with the motion of a body part (e.g., hand). Generally speaking, however, multiple points of attention co-exist and this represents a not trivial problem for most currently available models of visual attention. We hypothesize that combining our framework with a more refined model of visual attention (for a review of bottom-up (image-based) and top-down (task-dependent) models visual attention, see Ref. [34]), would actually show that temporal integration provides a partial clue of how to select from a set of candidate points of visual attention. In a sense, when the hand or the tool touches the object, visual, tactile and proprioceptive sensory information co-occur, giving an additional feature which can be used in the attentional decision-making process.

In order for our model to be more realistic, additional (possibly important) factors have to be included. For example, the model does not take into consideration the shape of the object or the appearance of the tool and its similarity with the limb (here, the hand). Both factors have been shown to influence body schema or image extension [19]. However, because we did not consider these factors, our model is able to deal with extended and non-extended limbs homogeneously regardless of their shape or appearance.

A further issue concerns the use of temporal integration as the only underlying mechanism. The key property of temporal integration of sensory information is that it holds independently of form and shape of the tool, e.g., a robot can reach for and

touch objects with its hand, with a small tool or with a big tool. It follows that any more advanced version of our model will also have to include temporal integration as one of its fundamental principles. In other words, the proposed model can be scaled up to more complex scenarios, maintaining the same underlying principle. Timing seems to be an important and invariant characteristic of tool-use in general. When touching an object using a stick or another tool (e.g., striking a nail with a hammer), there is typically a displacement between visual and tactile information. Despite the visual displacement (and thus the difficulty of integrating information spatially), the timing is identical, i.e., visual, tactile and proprioceptive activity co-occur, and information can be integrated temporally. Also related to the scalability issue is the inclusion of additional modalities. For instance, auditory signals such as a slapping sound could be used as a reinforcement signal for tactile sensation [35].

Before concluding the discussion, we reiterate that the main point here is that generalization from hand to tool is possible, and is based on previously acquired knowledge — in this case, the knowledge about the timing of visual and tactile information. We suggest that it is possible to generalize this idea to even more complex tool-use skills.

7. CONCLUSION

We proposed a novel model of tool-use-dependent body schema adaptation. Our model was inspired by neurophysiological findings, particularly a phenomenon known as 'body schema extension', which has been observed in monkeys and in humans. We tested our model by instantiating it in a robot whose task was to learn to use a simple tool to extend its reach. Our experimental results show that the model allows the robot to extend its body schema to incorporate an external (extra-corporeal, inanimate) object through the temporal integration of multisensory information (tactile and visual). We conclude that to learn to use tools through actor–environment interaction, body schema adaptation might actually be a good strategy. Another conclusion is that a plastic body schema is a necessary requirement for tool-use skills to emerge and that the approach in which a body schema is first constructed may have actual validity for achieving higher-level functions.

Our model is in line with the approach advocated by developmental robotics, but represents only the first step towards the acquisition of a truly adaptive robotic body schema. If a robot could acquire a body representation similar to ours, then it might actually also develop high-level cognitive functions such as imitation [1, 14] or spatial perception [2]. Future work will be mainly aimed at extending the model proposed in this paper to include imitative learning.

Acknowledgements

Parts of this study were supported by a Grant-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science.

REFERENCES

1. M. Merleau-Ponty, *Phénoménologie de la Perception*. Gallimard, Paris (1945).
2. V. Fogassi, G. Rizzolatti, L. Fadiga and V. Gallese, The space around us, *Science* **277** (5323), 190–191 (1997).
3. B. B. Beck, *Animal Tool Behavior: The Use and Manufacture of Tools by Animals*. Garland Press, New York (1980).
4. G. Berlucchi and S. Aglioti, The body in the brain: neural bases of corporeal awareness, *Trends Neurosci.* **20**, 560–564 (1997).
5. A. Iriki, M. Tanaka and Y. Iwamura, Coding of modified body schema during tool use by macaque postcentral neurones, *Cognitive Neurosci. Neuropsychol.* **7**, 2325–2330 (1996).
6. S. H. Johnson-Frey, The neural bases of complex tool use in humans, *Trends Cognitive Sci.* **8**, 71–78 (2004).
7. A. Maravita and A. Iriki, Tools for the body (schema), *Trends Cognitive Sci.* **8**, 79–86 (2004).
8. S. Gallagher, Body image and body schema: a conceptual clarification, *J. Mind Behav.* **7**, 541–554 (1986).
9. S. Gallagher, *How the Body Shapes the Mind*. Oxford University Press, Oxford (2005).
10. H. Head and G. Holmes, Sensory disturbances from cerebral lesions, *Brain* **34**, 102–245 (1911).
11. A. Maravita, C. Spence and J. Driver, Multisensory integration and the body schema: close to hand and within reach, *Curr. Biol.* **13**, R531–R539 (2003).
12. N. P. Holmes and C. Spence, The body schema and multisensory representation(s) of peripersonal space, *Cognitive Process.* **5**, 94–105 (2004).
13. P. Schilder, *The Image and Appearance of the Human Body: Studies in the Constructive Energies of the Psyche*. Kegan Paul, London (1935).
14. S. Gallagher and A. Meltzoff, The earliest sense of self and others: Merleau-ponty and recent developmental studies, *Philos. Psychol.* **9**, 213–236 (1996).
15. M. Lungarella, G. Metta, R. Pfeifer and G. Sandini, Developmental robotics: a survey, *Connect. Sci.* **15**, 151–190 (2003).
16. M. L. Simmel, The conditions of occurrence of phantom limbs, *Proc. Am. Philos. Soc.* **102**, 492–500 (1958).
17. M. L. Simmel, The absence of phantoms for congenitally missing limbs, *Am. J. Psychol.* **74**, 467–470 (1961).
18. V. S. Ramachandran and D. Rogers-Ramachandran, Synaesthesia in phantom limbs induced with mirrors, *Proc. R. Soc. Lond.* **263**, 377–386 (1996).
19. V. S. Ramachandran and S. Blakeslee, *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. Harper Collins, London (1998).
20. M. Spong and M. Vidyasagar, *Robot Dynamics and Control*. Wiley, New York (1989).
21. S. Ganapathy, Decomposition of transformation matrices for robot vision, in: *Proc. Int. Conf. on Robotics and Automation*, Atlanta, GA, Vol. 1, pp. 130–139 (1984).
22. Y. Nakagawa, H. G. Okuno and H. Kitano, Using vision to improve sound source separation, in: *Proc. Natl. Conf. on Artificial Intelligence*, Orlando, FL, pp. 768–775 (1999).
23. H. Ritter, T. Martinetz and K. Shulten, *Neural Computation and Self-Organizing Maps*. Addison-Wesley, New York (1992).
24. J. J. Craig, *Adaptive Control of Mechanical Manipulators*. Addison-Wesley, Reading, MA (1988).
25. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA (1998).
26. M. Kawato, Internal models of motor control and trajectory planning, *Curr. Opin. Neurobiol.* **9**, 718–727 (1999).
27. D. M. Wolpert, Z. Gharhramani and J. R. Flanagan, Perspectives and problems in motor learning, *Trends Cognitive Sci.* **5**, 487–494 (2001).

28. Y. Yoshikawa, K. Hosoda and M. Asada, Does the invariance in multi-modalities represent the body scheme? A case study with vision and proprioception, in: *Proc. 2nd Int. Symp. on Adaptive Motion of Animals and Machines*, Vol. SaP-II-1 (2003).
29. A. Stoytchev, Computational model for an extendable robot body schema, *Technical Report GIT-CC-03-44*, College of Computing, Georgia Institute of Technology (2003).
30. A. Stoytchev, Behavior-grounded representation of tool affordances, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Barcelona, pp. 3071–3076 (2005).
31. G. Metta and P. Fitzpatrick, Early integration of vision and manipulation, *Adaptive Behav.* **11**, 109–128 (2003).
32. M. Morita, Memory and learning of sequential patterns by nonmonotone neural networks, *Neural Networks* **9**, 1477–1489 (1996).
33. E. Oja, A simplified neuron model as a principal component analyzer, *J. Math. Biol.* **15**, 267–273 (1982).
34. L. Itti and C. Koch, Computational modelling of visual attention, *Nat. Rev. Neurosc.* **2**, 1–11 (2001).
35. S. Shimojo and L. Shams, Sensory modalities are not separate modalities: plasticity and interactions, *Curr. Opin. Neurobiol.* **11**, 505–509 (2001).
36. P. Haggard and D. M. Wolpert, in: *Higher-order Motor Disorders: From Neuroanatomy and Neurobiology to Clinical Neurology*, H. J. Freund, M. Jeannerod, M. Hallett and R. Leiguarda (Eds), pp. 261–271. Oxford University Press, Oxford (2005).

ABOUT THE AUTHORS

Cota Nabeshima received his BE and ME in Information Science and Technology from the University of Tokyo in 2004 and 2006, respectively. He is currently a PhD candidate at the Department of Mechano-Informatics. His research interests are interdisciplinary, and include adaptive robotics cognitive developmental science and embodied AI. His research goal is to devise computational strategies to make robots become more intelligent tool-users. He is a student member of the Robot Society Japan.

Max Lungarella received his PhD at the Artificial Intelligence Laboratory of the University of Zurich, Switzerland, in 2004. From 2002 to 2004, he was an Invited Researcher at the Neuroscience Research Institute of the National Institute of Advanced Science and Technology, Japan. Currently, he is a Postdoctoral Researcher at the Department of Mechano-Informatics, University of Tokyo. He has been involved in various projects related to artificial intelligence, robotics and electrical engineering. His research interests include, but are not limited to, embodied artificial intelligence, developmental robotics, information theory, motor control, and electronics for perception and action.

Yasuo Kuniyoshi is a Professor at the Deparment of Mechano-Informatics, University of Tokyo, Japan. He received his ME and PhD degrees from the University of Tokyo in 1988 and 1991, respectively. From 1991 to 2000, he was first a Research Scientist and then a Senior Research Scientist at the Electrotechnical Laboratory, AIST, Japan. From 1996 to 1997, he was a Visiting Scholar at the MIT AI Lab. In 2001, he joined the Department of Mechano-Informatics, University of Tokyo. He is the author of over 200 technical publications, editorials and books. He received an Outstanding Paper Award from

the International Joint Conference on Artificial Intelligence, a Best Paper Award from the Robotics Society of Japan and the Sato Memorial Award for Intelligent Robotics-Research. His research interests include emergence and development of embodied cognition, human action understanding systems, and humanoid robots. He is a member of IEEE, the Robotics Society of Japan, the Japanese Society for Artificial Intelligence and the Japanese Society of Baby Science.