Ms. 2672

Is there something out there? Inferring space from sensorimotor dependencies.

D. Philipona[*], J.K. O'Regan[†], J.-P. Nadal[‡]

February 17, 2003

Abstract

This paper suggests that in biological organisms, the perceived structure of reality, in particular the notions of body, environment, space, object, and attribute, could be a consequence of an effort on the part of brains to account for the dependency between their inputs and their outputs in terms of a small number of parameters. To validate this idea, a procedure is demonstrated whereby the brain of an organism with arbitrary input and output connectivity can deduce the dimensionality of the rigid group of the space underlying its input-output relationship, that is the dimension of what the organism will call physical space.

1 Introduction

The brain sits inside the cranial cavity monitoring the neural signals that come into it, and go out of it. From this processing emerge the notions of self, of outside space, of objects within that space, and of object attributes like color, luminosity, temperature. Even simple organisms that have little or no cognitive abilities clearly possess these concepts at least implicitly, since they show spatially adapted

[*]Sony CSL, 6 rue Amyot, 75005 Paris, France, david.philipona@polytechnique.org

[†]Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes, 71, avenue Edouard Vaillant, 92774 Boulogne-Billancourt Cedex, France, http://nivea.psycho.univ-paris5.fr

[‡]Laboratoire de Physique Statistique, Ecole Normale Supérieure, rue Lhomond, 75231 Paris Cedex 05, France, http://www.lps.ens.fr/∼nadal

behavior like locomotion, navigation, grasping and discrimination of different objects.

How is this possible? What kind of algorithms must be at work inside biological brains for these notions to be extracted from the neural activity in a mass of unlabelled nerve fibres? Do brains have this capacity because phylogeny has yielded a brain structure which is specially adapted to understanding the notion of space?

Here we investigate a more radical hypothesis, namely the possibility that what brains do is to continuously calculate statistics on their neural activity in an attempt to characterize it *using a small number of parameters*. We take the extreme case where the brain has absolutely no a priori information about outside physical space (whether it exists at all, whether it has a metric, whether it is euclidean, how many dimensions it possesses). We assume that the linkage between motor commands and the resultant motions of the organism is unknown to the brain and totally arbitrary. We further assume that the brain has no information about what nerve pathways correspond to sensors sensing internal or external states.

We show that there is a simple procedure which a brain can use to arrive at a distinction between the body, which it can control, and the outside world, which it cannot completely control. There is a simple algorithm that leads to a characterization of the number of variables necessary to describe the organism's body (insofar as the body states affect incoming sensory information).

Further, and most interesting, we show that the brain can deduce the dimensionality of outside physical space, and the number of additional non-spatial parameters needed to describe the attributes of objects or entities within it.

Our hypothesis is therefore that what biological organisms perceive as being the limits of their bodies, as well as the geometry and dimensionality of space outside them, are deducible, without any a priori knowledge, from the laws linking the brain's inputs and outputs. The approach we are taking derives from the basic idea that the basis of sensory experience consists in extracting and exercising laws of sensorimotor dependencies (O'Regan & Noë, 2001).

2 A simple organism

Let us imagine a simple organism consisting of an articulated arm fixed to the ground. At the end of each of its fingers it has an eye composed of a number of light-sensitive sensors. Imagine in addition that the organism has proprioceptive devices that signal the position of the different parts of the arm. The environment,

we shall suppose, consists in a set of lights. The signals provided by the sensors are transmitted to a brain that controls effectors that move the arm.

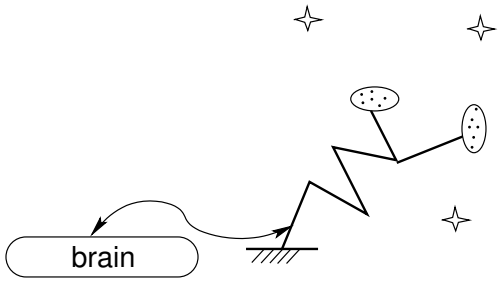

Figure 1: A simple organism consisting of an articulated arm with two "fingers" and a composite "eye" mounted on each.

Let us suppose that the brain has no a priori knowledge about the body that it is connected to, and that this body is the sole connection it has to the outside world. What comprehension of this "exterior" can the brain attain, and how can it be obtained?

Certainly the simplest thing the brain can do is to start by trying various motor commands, at random. Gradually the brain will notice that it can make a distinction among the mass of sensory inputs that it receives: It will notice that certain inputs, or certain combinations of inputs, always react the same way to motor commands, while the other inputs show only partial, unsystematic relations to motor commands. What is the natural conclusion that the organism can deduce from this fact? It is the fact that its universe can be separated into a part which the organism can completely control, and a part that the organism can only partially control.

We shall call the first part, over which it has complete control: the organism's *body*, and the second part the organism's *environment*. We shall call the first type of inputs *proprioceptive* and the others *exteroceptive*[1]. We shall say the *body is stationary* when proprioception is constant, and we shall say the *environment is*

[1]We follow the terminology used in Kandel, Schwartz, & Jessell, 2000, and stress the fact that exteroceptive sensors are not only sensitive to changes of the environment but to motion of the body as well, while proprioceptive are sensitive to changes of the body *only* (which is more restrictive that the usual use of this term). Also, it should be noted that this distinction arises gradually: certain inputs which might at first seem completely determined by motor commands will later turn out in fact only to be partially determined by them.

stationary when exteroception is constant[2]. Note that since the organism is totally naive about its environment (and even about the fact that there is such a thing at all as an environment), it has no choice but to *define* these notions.

The brain can now attempt to understand its environment. Changes that occur in exteroceptive sensors when the body is stationary can be taken to derive from the environment. We shall assume that the brain attempts to account for these changes, which are defined in the very high dimensional space of the number of exteroceptors, in terms of a much smaller number of parameters. For example, in the case of the articulated arm, there are 40 photoreceptors, but their outputs are completely determined by a much smaller number of parameters, namely the positions of the three lights in the environment. We shall call the vector of this reduced parameter set a *representation of the state of the environment*.

The brain can also issue motor commands. When it does this sufficiently quickly it can assume that most of the resulting changes in its exteroceptive input will be due to its own motion and not to motion of the environment. Again a low-dimensional parameter set can be extracted from the exteroceptive changes that occur in this case, and we shall call the resulting vector a *representation of the exteroceptive body*.

The representation of the exteroceptive body is richer than the representation of the proprioceptive body in the sense that it provides a way for the organism to discover the *relationship* between its body and the state of the environment. Importantly, the organism can note that certain exteroceptive changes caused by motion in those body parts that bear sensors, can be *compensated* by particular environmental changes. For example, in the example of the articulated arm, when the set of photoreceptors is displaced rigidly by an arm motion, there is a corresponding inverse rigid displacement of the environmental lights which can be made, which will cause exteroception to return to its original value. Note however that such compensations are *possibilities* that need never actually occur for our algorithm to work – it is unlikely that the environment would ever actually move in this particular rigid fashion. Note also that the notion of compensability derives in an intrinsic fashion from the fact that the response of a single set of sensors, namely the exteroceptive ones, varies as a function of two distinct sources of variation: body changes and environmental changes. Proprioceptive sensors, since they are only sensitive to a single source of variation, namely body changes, do not provide information about this kind of compensability.

[2]This definition is compatible with sensors sensitive to derivatives of position, since for example if both velocity and position are constant, then velocity must be zero.

The notion of compensability, arising from the confrontation of two distinct sources of change within a single set of inputs, has fundamental implications for the genesis of the notion of space in an organism:

1. Compensability defines a class of body movements with a particular structure: if the body makes such a movement, then the movement bringing the body back to its original position is also compensable. If the body makes two successive compensable movements, then the resulting global movement is also compensable. In other words, this class of movements has the mathematical structure of a group, whose identity element is stationarity. We shall call an element of this group a *compensable transformation* of the exteroceptive body. In the same way, we can define the group of compensable transformation of the environment and the group of compensated transformations of the body-environment system. By their definitions, these three groups are very closely related to each others.

2. Compensability implies that there is something in common between certain body movements and certain environmental movements. It now (but not before) becomes possible to say that body and environment are immersed in a single entity which we call *space*. Note that it is through the use of a sensorimotor approach that we have attained the notion of space: there can be no notion of compensability when we only make passive observations. It is through actions that arises the idea of an environment distinct from the body, and through actions that is born the idea of a common medium through which to describe the body and the environment. In a similar vein, Poincaré pointed out the radical incommensurability between sensations originating in different sensory modalities, unified only by common motor acts needed to reach their sources (Poincaré, 1895, 1902).

Certain compensated transformations have an additional, special, structure. They form a non-trivial subgroup, in the sense that the sensory consequences of two successive transformations of that particular kind may depend on the order in which they have been performed. On the other hand, other compensated transformations can be applied before or after any other compensated transformations (including the previously mentioned ones) without changing anything from the sensory point of view. We will call the first transformations *rigid transformations* and the others *attribute transformations*. Indeed, the first group of transformations defines space, and the other transformations define changes over non-spatial attributes.

The terminology of "rigid" transformations comes from the fact that these transformations are exactly related to those spatial transformations of the body-environment system leaving the sensory inputs fixed. Thus, any metric defined over representations of space based on the sensory inputs only will have to be constant over these latter changes.

Since our organism is assumed to be totally ignorant about itself and its surroundings, it has no choice but to *define* rigidity — the same problem was encountered for the notion of stationarity. This means that if, for example, the physical laws obeyed by light changed according to the positions of the sensors, there would be a difference with the euclidean notion of rigidity. But physicists themselves encounter the same problem in their conception of space, since they must also assume as *given* the notion of a rigid measuring rod.

We have attained the non-trivial conclusion that without observing its exteroceptive body "from the outside", the brain can deduce that there exist transformations of the body-environment system that its exteroception is not sensitive to. These compensated transformations provide the brain with what it needs to ground its understanding of geometry.

It is worthwhile to point out the relation of our approach with that of Droulez & Berthoz, 1991; Blakemore, Frith, & Wolpert, 2001; Blakemore, Wolpert, & Frith, 2002, where it is proposed that the cerebellum attempts to subtract a prediction of the sensory consequence of its movements from the actually observed consequences in order to deduce the changes that must have occurred in the environment. The similarity lies in the very general idea of analyzing the sensorial consequences of a movement of the body alone to understand the changes of the environment. This is a classical idea today. But it usually relies on a kind of platonic a priori about the existence of space and assumes that the role of the brain is to map its sensory inputs to some kind of objective archetype of the world, and try to understand its sensations in relation to this abstract world. We used terms such as "representations of the state of the exteroceptive body" to describe what *we* (or the brain) conceive this world to be, without any relationship with an a priori model.

3 Mathematical sketch

In order to make the preceding discussion more precise, and in order to derive a simple, neuronally plausible algorithm, we here present a sketch of a mathematical

formalization. We illustrate only the essential aspects of our approach to show in a few steps how it is possible to deduce the dimension of the manifold of rigid transformations of outside "physical" space. The Appendix provides suggestions for a more realistic implementation, and a second paper will show, beyond the discovery of dimensions, how the group structure of these rigid transformations can be accessed and used.

We think that the problem we want to answer is precisely the problem addressed in differential geometry. Indeed, a usual way of introducing the aim of differential geometry is transparently summarized by saying "to consider $\mathcal{S}$ as a manifold means that one is interested in investigating those properties of $\mathcal{S}$ that are invariant under coordinate transformations" (Amari & Nagaoka, 1993). If we think of the sensorimotor system as a highly redundant parameterization system to move in this manifold and make observations on it, then the goals are strikingly identical. The structure of the world consists in those properties that will be imposed on any sensorimotor system, the rest is an artefact of the specificities of bodies. It is in this spirit that we use the language of differential geometry, with the conviction it can provide insights.

Consider an environment whose set of all states E is a manifold $\mathcal{E}$ of dimension e. Suppose the set of all observed sensory inputs S is a manifold $\mathcal{S}$ of dimension s, and the set of all possible outputs M is a manifold $\mathcal{M}$ of dimension m. Finally, suppose the environment imposes a 'smooth enough' (meaning we will consider the problem only in a region excluding the singularities of ψ) relation between sensory signals S and motor commands M:

$$S = \psi(M, E) \tag{1}$$

Note that with our definitions we have $\mathcal{S} = \psi(\mathcal{M} \times \mathcal{E})$.

Though the mathematics does not require it, to facilitate the discussion below we shall consider the case where the manifolds involved are embedded in finite vector spaces, and that (S, M) is a vector whose coordinates reflect the neural activities of the sensory and motor neurons at some time t. Making this choice has the consequence that the sensory inputs will have to be determined instantaneously by the motor output and the environment states. This is thus inappropriate for cases where the sensory inputs are determined by the integration over time of motor commands, as is the case when we control the rotation speed of the wheels of a robot, for instance. Our example would apply instead to a case where the neural motor activity would be enough to infer muscle elongation and thus the position of the body. However we wish to emphasize that this choice of instantaneous relation between sensory and motor activities is not necessary from the

mathematics and is mainly used to make the presentation below more intuitive. Furthermore, it will allow us to identify proprioception very easily.

Indeed, because we argue that geometry arises through compensability, and that proprioceptive inputs cannot be compensated by any change of the environment, the first thing we want to do is to distinguish proprioceptive from exteroceptive inputs. This can easily be done in the previous model by locating inputs which do not change when a given definite motor order is issued and maintained. From now on we shall be concerned only with exteroceptive inputs.

Following the method frequently adopted in motor control (Atkeson & Schaal, 1995; Vijayakumar & Schaal, 2000; Baraduc, Guigon, & Burnod, 2001, although we are *not* using a motor control approach $M = \varphi(S)$, but an approach *on the contrary* based on the observation of sensory consequences of motor commands) and the standard mathematical approach for dealing with manifolds, we shall focus on the tangent space $\{dS\}$ of $\mathcal{S}$ at some point $S_0 = \psi(M_0, E_0)$.

This is allowed by the smoothness of ψ, and it is important to note that this provides an exact theoretical ground for the considerations below, while the robustness of the local linear approximation we will make in practice is a different issue. Our aim was to show that the dimension of the rigid group of space is accessible through sensory inputs, and to demonstrate the conceptual framework needed to access it. The actual implementation we used had no other purpose than illustration, and the question of its robustness for practical robotic applications has no incidence on our more theoretical point. The simple, neuronally plausible (Sejnowski, 1977; Oja, 1982; Sanger, 1989) tool we used to estimate these dimensions is Principal Component Analysis (PCA), but any other method of dimension estimation would have been suitable.

This being said, two natural subspaces can be identified in $\{dS\}$: the vector subspace $\{dS\}_{dE=0}$ of sensory input variations due to a motor change only, and the vector subspace $\{dS\}_{dM=0}$ of sensory input variations due to an environment change only. Since we have:

$$dS = \frac{\partial \psi}{\partial M}|_{(M_0,E_0)} \cdot dM + \frac{\partial \psi}{\partial E}|_{(M_0,E_0)} \cdot dE \tag{2}$$

we can remark that

$$\{dS\} = \{dS\}_{dM=0} + \{dS\}_{dE=0} \tag{3}$$

What are the manifolds to which these subspaces are tangent? Starting from (M_0, E_0), we can consider the sensory inputs obtained through variations of M only (E_0-section $\equiv \psi(E_0, \mathcal{M})$), and sensory inputs obtained through variations of

E only (M_0-section $\equiv \psi(\mathcal{E}, M_0)$) (see figure 2). $\{dS\}_{dE=0}$ and $\{dS\}_{dM=0}$ are the tangent spaces at the point S_0 of these manifolds, and the fact that their vectorial sum is the overall tangent space of $\mathcal{S}$ means, by the definition of transversality, that the two sections are transverse (see for instance Laudenbach, 1999 for a basic description of transversality in submanifolds of $\mathcal{R}^n$.). We will call $\mathcal{C}(M_0, E_0)$ their intersection, which is thus a manifold as well.

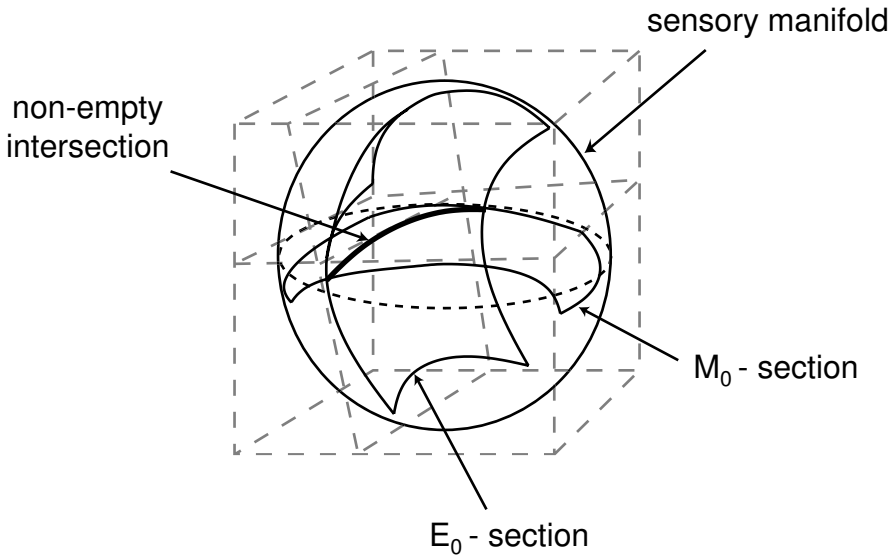

Figure 2: The sensory manifold in the neighborhood of S_0, the E_0 and M_0-sections (see text). These two manifolds are transverse, and their intersection is the manifold of the sensory inputs accessible either through motion of the exteroceptive body or motion of the environment.

When the body is stationary, the dimension of $\{dS\}_{dM=0}$ gives the number e of variables necessary for a local description of the environment. When the environment is stationary, the dimension of $\{dS\}_{dE=0}$ gives the number $p \leq m$ of variables necessary to explain the variations in exteroceptive signals due to body motions, that is to say, the number of variables describing the exteroceptive body. When M and E both vary, $\{dS\}$ is not of dimension $p + e$ because the vector spaces generated by $\frac{\partial \psi}{\partial M}$ and $\frac{\partial \psi}{\partial E}$ do not necessarily have null intersection. Certain exteroceptive changes can be obtained equally either from dE or from dM, as it is the case when we move along $\mathcal{C}(M_0, E_0)$. This reduces the dimensionality of dS

compared to the case where it is possible to identify the origin of the change unambiguously. When sensory changes from dE and dM mutually compensate for one another we shall say that there has been a compensated infinitesimal movement of the body-environment system, and we shall say that the corresponding infinitesimal movement of the body (or the environment) is compensable (thus, implicitly, the term of compensated movement means a change of body-environment, while compensable movement means a change of either one alone). For simplicity we will drop the word "infinitesimal", but it should be always kept in mind since we will mostly be dealing with the tangent spaces. This is justified because it is a basic result that the dimension of a tangent space is the dimension of the underlying manifold.

Since the exteroceptive body representation and the environment representation imply parameterizations of $\{dS\}_{dM=0}$ and $\{dS\}_{dE=0}$, there is a natural one-to-one correspondence between the set of compensated movements, and the set of compensable movements of the body. The basic idea is that, given a compensable movement of the body, only the environment movement yielding the opposite sensory consequence will, when taken together, yield a compensated movement of the body-environment system[3].

We will now search for the relationships between the dimensions of all the entities we have defined. We have shown that the dimension of the space of compensated movements is the same as the dimension of the space of compensable body movements, and this dimension is accessible because there is a trivial one-to-one mapping of this space to $T = \{dS\}_{dM=0} \cap \{dS\}_{dE=0}$. Indeed, $dS_{dM=0}$ is compensable means:

$\exists\ dS_{dE=0}$ such that $dS_{dM=0}+dS_{dE=0} = 0 \Leftrightarrow dS_{dM=0} = -dS_{dE=0} \Leftrightarrow dS_{dM=0} \in T$

Thus the space of the compensated movements has the dimension of T. But since

$$\begin{aligned} \dim\{dS\}_{dM=0} + \{dS\}_{dE=0} &= \\ \dim\{dS\}_{dM=0} &+ \dim\{dS\}_{dE=0} - \dim\{dS\}_{dM=0} \cap \{dS\}_{dE=0} \end{aligned}$$

we finally have, with the use of equation 3:

$$d = p + e - b \tag{4}$$

[3]This is enough for our purpose, but (for the extension of this work) it is important to understand that this can be extended to non-infinitesimal movements by integration: given a compensable movement of the body (i.e. a curve tangentially compensable at any time), we can construct the movement of the environment whose tangent sensory change will cancel at any time the tangent sensory change yielded by the compensable movement. This is symmetrically true for compensable movements of the environment.

where $b = \dim\{dS\}$ and $d = \dim T = \dim \mathcal{C}(M_0, E_0)$.

We have consequently deduced the dimension of the manifold of the compensated transformations of the world. Note that this is actually more fundamental than simply the number of variables needed to describe this world: the manifold of compensable transformations is the operational aspect of the relation to the world that an organism will be most interested in.

In a subsequent, more technical paper (in preparation) we demonstrate how the group structure of the underlying transformations can be accessed and defined through $\mathcal{C}(M_0, E_0)$. This will provide a way for the naive organism to understand its body and produce at will particular rigid transformations of its exteroceptive body, such as translations or rotations for instance. The dimension of the rigid group, and the dimension of space will be found through the study of these transformations. For now, if we make the additional assumption that the organism can choose to perform only spatial movements, then we can say that we have accessed the dimension of the rigid group defining the space the organism is embedded in.

A last point must be made concerning the distinction between proprioceptive and exteroceptive sensory inputs. In the preceding mathematical discussion we have assumed we are dealing only with exteroception, and that proprioception has been set aside. It might be thought that this is an unnecessary step, since in equation 4 proprioceptive dimensions surely would cancel out, since they would contribute equally to p and b. However, this is in general false. If we consider a case where the representation of the exteroceptive body intersects with the representation of the proprioceptive body, then describing this intersection will, in a moving environment, require two times the set of variables required in a fixed environment. Indeed, in a moving environment exteroception and proprioception will be unlinked.

4 Experiments

4.1 Method

We here present the results of a simulation of the example of the articulated arm described above. We additionally describe simulations for two further cases with interesting modifications of the organism. The details of the three experiments can be found in the Appendix. A summary of the results is presented in Table 1. It should be stressed that the same kind of simulation could be done for any other arbitrary kind of device with sensory inputs and motor outputs.

Caracteristics	Organism 1	Organism 2	Organism 3
Dim. of motor commands	40	100	100
Dim. of exteroceptive inputs	40	80	80
Number of eyes	2	4	4
Diaphragms	none	reflex	controlled
Number of lights	3	5	5
Light luminance	fixed	variable	variable
Dim. found for body (p)	12	24	28
Dim. found for environment (e)	9	20	20
Dim. found for both (b)	15	38	41
Deduced dim. of rigid group (d)	6	6	7

Table 1: Summary of the three experiments. Proprioception does not play a role in the calculation and so is not shown in the table. The estimations given here are obtained from graphs 3.c and 4. In Organism 3 the group of compensated transformations is different from the orthogonal euclidean group because the organism has control over a non-spatial aspect of its body, namely the diaphragm aperture.

In the first experiment the arm had four joints and two eyes, and the environment consisted of three lights. Each eye consisted of a composite "retina" with 20 omnidirectionally, that is to say not directional, sensitive photo-sensors mounted rigidly on a small flat surface, attached to the end of a "finger", one for each eye.

Each joint had 4 proprioceptive sensors whose output depended on the position of the joint, according to a fixed, randomly assigned law. The orientation of the eyes provided no proprioception.

The motor command moving the device was a 40-dimensional vector, which was converted by a fixed random function to the 12 values that determined the 3D spatial coordinates of the surfaces holding the two eyes and their orientations.

All these particular choices were arbitrary: the purpose was merely to simulate a complicated sensorimotor relation that was unknown to the brain and had the property that the number of dimensions of the motor commands and of the sensory inputs should be high compared to the number of degrees of freedom of the physical system.

In the second experiment we considered a more complex device with an arm having ten joints, bearing four eyes. Each eye had a diaphragm or attenuator with an automatic "pupil-reflex" that reduced light input to it in such a way that total illumination for the eye was constant. There were five light sources in the environ-

ment, and we now allowed their intensity to vary. The dimensionality of the motor command and sensory input was also increased respectively to 100 (determining the 24 degrees of freedom of the four eyes, each having three positional and three orientational degrees of freedom) and 120 (determined by the 20 photosensors on each of the four eyes plus 40 proprioceptors).

Again the purpose was to show that the complexity of the sensorimotor coupling was not a theoretical obstacle to our approach, neither were non-spatial body changes like the pupil reflex, nor non-spatial changes in the environment, like light intensity.

The third experiment was identical to the second except that now we made what we shall see below is a theoretically very important modification: the diaphragms were now controlled by the organism instead of being determined automatically by the total illumination.

To do the simulation we went through the four-stage procedure described in the previous section.

1. Proprioceptive input was separated from exteroceptive input by noting that proprioceptive input remains silent when no motor commands are given, whereas exteroceptive input changes because of environmental change.

2. We estimated the number of parameters needed to describe the variation in the exteroceptive inputs when *only the environment* changes. The algorithm issues no motor commands, and simply calculates the covariance matrix of the observed environment-induced variations in sensory inputs. The dimension estimation is done by considering the eigenvalues of this covariance matrix. The eigenvalues λ_i should fall into two classes, a class with values all equal to zero, and a class with non-zero values. We separated the two classes by a clustering method (see Appendix). The number of non-zero eigenvalues was taken as the number of dimensions.

3. We estimated the number of parameters needed to describe the variation in the exteroceptive inputs when *only the body* moved. The environment is kept fixed and the algorithm gives random motor commands (see Appendix). We observe the covariance matrix of the resulting changes and estimate the dimension from the number of non-zero eigenvalues in the same way as before.

4. We estimate the number of parameters needed to describe the changes in exteroceptive inputs when *both the body and the environment* change. The

environment is changed at random, and the organism gives random motor commands. The number of non-zero eigenvalues of the covariance matrix is obtained as before.

4.2 Results for Organism 1

We see that despite the high dimensionality of the outputs, the exteroceptive, and the proprioceptive inputs (40, 40, 20, respectively), the brain is able to:

1. distinguish the 20 proprioceptive and 40 exteroceptive sensor inputs (Figure 3.a).

2. determine that only 12 parameters are necessary to represent its exteroceptive body — this corresponds to the 3 position and 3 orientation dimensions for each of the two eyes (Figures 3.b and 3.c).

3. determine that 9 parameters are necessary to describe the environment — corresponding to 3 spatial dimensions for each of the 3 lights (Figures 3.b and 3.c).

These results would seem to lead to the conclusion that the brain should observe $12 + 9 = 21$ degrees of freedom in the exteroceptive inputs when it lets body and environment move simultaneously, but instead only 15 are found (3.b and 3.c). The brain thus concludes that there is a group of compensated movements of dimension $21 - 15 = 6$, that we know to be the Lie group of orthogonal transformations (3 translation and 3 rotations)[4].

4.3 Results for Organisms 2 and 3

In the second experiment, the algorithm deduced that 24 variables were needed to characterize the exteroceptive body, 20 were needed for the environment, and 38 were needed to characterize the exteroceptive inputs when both environment and body moved simultaneously (see Appendix). The group of compensated movements constituted by the brain therefore remained coherent with

[4] An elementary presentation of Lie groups can be found in Burke, 1985. The orthogonal transformations are clearly a subset of the set of compensated movements, and the equality of dimensions guarantees equality of the groups.The idea that sensory information should be invariant under the Lie group of orthogonal transformations is a useful concept in artificial vision (Van Gool, Moons, Pauwels, & Oosterlinck, 1995) and image analysis (Rao & Ruderman, 1999).

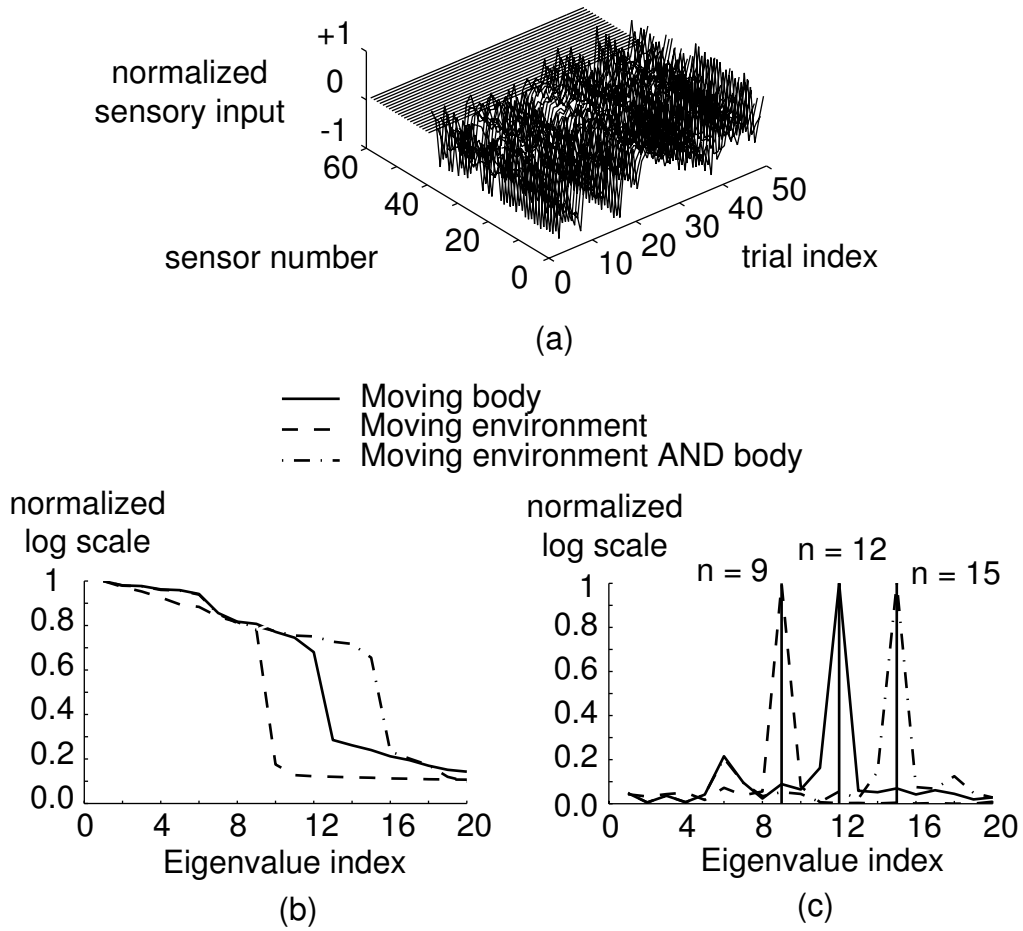

Figure 3: a) Changes in the different inputs when no motor commands are given and the environment changes. The first 40 sensors depend on the environment (exteroceptive) and the 20 last ones do not (proprioceptive). b) Eigenvalues of the covariance matrices in the three cases described in the text (normalized log scale: logarithm of the eigenvalues renormalized to lie in the range [0,1]). c) Ratio of eigenvalue i to eigenvalue $i+1$, where i is the eigenvalue index. The dimensions of the tangent spaces are taken to be equal to the number of significantly non-zero eigenvalues, and the maximum of this ratio indicates the biggest change in the order of magnitude of the eigenvalues (see Appendix).

the usual euclidean group, since the brain again arrived at a group of dimension $20 + 24 - 38 = 6$ (which is the same Lie group as before). We see that the notion of space constituted by the brain is insensitive to major variations in the complexity of the physical structure of the device and to the details of its sensorimotor coupling.

The third experiment involved the same organism as the preceding example, but the aperture of the diaphragms was now controlled by the brain, instead of being determined by an automatic reflex. This time the brain determined that 28 variables were needed to characterize the exteroceptive body. 20 were needed for the environment and 41 for the exteroceptive inputs when both environment and body moved. Luminance in this example is a compensable variable and we indeed found that we now have $28 + 20 - 41 = 7$ compensated variables instead of 6.

The group structure, and more precisely commutativity of the transformation involving only luminance variations with any other transformation, still theoretically allow for the distinction between spatial transformations and attribute changes. But it is worthwhile to recall that Poincaré thought of geometry in terms of *voluntary* body motions. It would seem that he considered that only spatial changes were subject to voluntary control. If space is thus defined by the voluntary accessible set of sensory inputs, then the control of our diaphragms would surely have led us to a different conception of space than the one we have. If we define space through the properties of commutativity of transformations operating on it, which is the hypothesis we favor, then the diaphragms are of no importance.

A second aspect of this third experiment should not be dismissed, which concerns the notion of environmental attributes. By allowing luminance to vary independently in the five lights, one might have expected that the brain would find 5 additional variables as compared to the case where luminance is fixed. However the algorithm deduces that only a single additional compensable dimension need be introduced. This is because compensability of luminance with motions of the body-environment configuration is possible only in the case where all five lights vary simultaneously in luminance by a common factor. Thus we see that the algorithm has extracted the presence of an additional feature dimension, namely luminance, independently of the number of lights that are present.

We wish to stress for the last time that the simulations presented here are considered simply as illustrations of our approach. We claim nothing about the optimality or robustness of the methods we have used: there is an active literature on the practical problems involved in estimating dimensionality and, more interesting, on the parametrization of manifolds (Tenenbaum, 1998; Roweis & Saul,

2000; Tenenbaum, Silva, & Langford, 2000; Verbeek, Vlassis, & Kröse, 2002). Our purpose here was to show how, when taken in conjunction with a sensorimotor rather than merely sensory approach, such tools could be used by an organism to generate the notion of space.

5 Conclusion

We have shown that thanks to a sensorimotor approach, the notion of rigid group of space, where space is an abstract frame in which the brain has good reasons to embed its body and its environment, is accessible to an organism without any a priori knowledge. This shows why in robotics and neurobiology it may be fundamental to take sensory and motor information into account *simultaneously*. Indeed, we have shown that doing so provides a meaningful way of reducing the dimensionality of the problem of interacting with the environment. This is directly of interest for roboticists concerned with unsupervised learning. It is indirectly of interest for neurobiologists and cognitive scientists, since it favors the idea suggested in O'Regan & Noë, 2001 that the brain should be studied from a sensorimotor approach rather than a stimulus based or motor control approach: until now it has never been shown how such an approach could access the notion of space, while it is a notion we all know to be present in our brains.

It must be noted that our approach represents a basically different approach than the dimension reduction techniques that are usually employed, for instance in pattern classification tasks for passive observers. Consider for example Roweis & Saul, 2000, which does dimension reduction on a set of faces with different expressions. In this situation, the variables that can be used to parameterize facial expressions will have no simple relation to the dimension of the space in which the faces are embedded. If such a technique is applied to samples of faces which can be both rigidly rotated and which can change expression, the parameters determining these two types of change would not be differentiated. On the other hand, by the use we are suggesting here of an active observer and the notion of compensated variables, it is possible for our algorithm to make a principled distinction between rigid deformations like rotations and non-rigid variations like facial expression changes. Our approach induces a difference between simple feature spaces, that do not distinguish attributes and geometry, and structured geometrical spaces.

Finally, perhaps the main interest of our approach is philosophical. Spatial reasoning is the basis of our comprehension of the world and our abstraction abilities, so much so that the nature of space itself has been considered to preclude

scientific investigation. Ever since the revolution of non-euclidean geometry more than a century ago, our euclidean intuition of space has been considered to derive from our experience with the world, and not from a mental a priori. But then if our notions of geometry derive from our relation with the world, to what extent do these notions depend on the world, and to what extent do they depend on the way our sensorimotor systems are constructed? Could it be that the intuition of three-dimensional euclidean space is a consequence of the structure of humans' sensorimotor systems rather than corresponding to "real" properties of the outside world?

To answer this question, we would have to know what we mean by "real" properties of the world, even though we obtain information about it only through our bodies and our sensors. This is problematic, since to understand the world we must know about how our bodies and sensors are constructed. But how can we formulate such knowledge without presupposing a world within which the body and its sensors are embedded, and whose concepts we use to describe those very bodies and sensors? It is because of this fundamental problem that we think it is fruitful to develop an abstract approach such as ours, in which the details of the workings of sensors and effectors play no role.

Our approach has shown that one reasonable deduction that brains can derive from the sensorimotor constraints they are exposed to is the notion of compensated movements. A brain can naturally infer that the structure of the derived compensated transformations of the world will be coherent with transformations derived by any other brain with the same sensorimotor capabilities, but different details of implementation. Indeed, since the compensated group is precisely those changes of the body-environment system leaving perception unchanged, if there is a one-to-one smooth mapping from the perception of one organism to the perception of the other, then they will end up with the same rigid group. The subsequent consensual universality of this group makes it a good candidate for the status of physical reality. We believe that it may be this which lies at the basis of our notion of the orthogonal group of geometrical space.

6 Acknowledgements

We would like to thank Olivier Coenen for his support and remarks, Jim Clark, Jacques Droulez, Daniel Bennequin, Alexandre Pouget, Jean Petitot, Patrick Gros and Christoph Zetzsche for their enlightening observations. This work was partly founded by the european project IST-2001-35271.

References

Amari, S.-I., & Nagaoka, H. (1993). *Methods of information geometry.* Oxford University Press.

Atkeson, C. G., & Schaal, S. (1995). Memory-based neural networks for robot learning. *Neurocomputing, 9*, 243-269.

Baraduc, P., Guigon, E., & Burnod, Y. (2001). Recoding arm position to learn visuomotor transformations. *Cerebral Cortex, 11*(10), 906-917.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport, 12*(9), 1879-84.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn Sci, 6*(6), 237-242.

Burke, W. L. (1985). *Applied differential geometry.* Cambridge University Press.

Droulez, J., & Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Sciences, 88*, 9653-9657.

Kandel, E. R., Schwartz, J., & Jessell, T. M. (2000). *Principles of neural science.* Mc Graw Hill.

Laudenbach, F. (1999). *Calcul différentiel et intégral.* Ecole Polytechnique.

Matarié, A. F. M. J., & Jenkins, O. C. (2000). Automated derivation of primitives for movement classification. *International Conference on Humanoid Robotics*.

Minka, T. P. (2000). Automatic choice of dimensionality for PCA. *NIPS 2000*, 598-604.

Oja, E. (1982). A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology, 15*, 267-273.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 883-917.

Poincaré, H. (1895). L'espace et la géométrie. *Revue de Métaphysique et de Morale, 3*, 630-646.

Poincaré, H. (1902). *La science et l'hypothèse.* Flammarion.

Rao, R. P. N., & Ruderman, D. L. (1999). Learning Lie groups for invariant visual perception. *Advances in Neural Info Processing Systems, 11*, 810-816.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*, 2323-2326.

Sanger, T. (1989). *Optimal unsupervised learning in feedforward neural networks* (Tech. Rep.). AI Lab MIT.

Sejnowski, T. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology, 4*, 303-321.

Tenenbaum, J. B. (1998). Mapping a manifold of perceptual observations. *NIPS-10*.

Tenenbaum, J. B., Silva, V. de, & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290*, 2319-2323.

Van Gool, L., Moons, T., Pauwels, E., & Oosterlinck, A. (1995). Vision and Lie's approach to invariance. *Image and Vision Computing, 13*(4), 259-277.

Verbeek, J. J., Vlassis, N., & Kröse, B. (2002). *Coordinating Mixtures of Probabilistic Principal Component Analyzers* (Tech. Rep.). The Netherlands: Computer Science Institute, University of Amsterdam. (IAS-UVA-02-01)

Vijayakumar, S., & Schaal, S. (2000). Fast and efficient incremental learning for high-dimensional motion systems. *International Conference on Robotics and Automation (ICRA2000)*, 1894-1899.

7 Appendix

7.1 Experimental Details

The motor commands were simulated in the following way:

$$\begin{aligned}(Q, P, a) &= \sigma(W_1 \cdot \sigma(W_2 \cdot M - \mu_2) - \mu_1) \\ L &= \sigma(V_1 \cdot \sigma(V_2 \cdot E - \nu_2) - \nu_1) \\ S^e_{i,k} &= d_i \sum_j \frac{\theta_j}{||P_i + Rot(a^\theta_i, a^\varphi_i, a^\psi_i) \cdot C_{i,k} - L_j||^2} \\ S^p_i &= \sigma(U_1 \cdot \sigma(U_2 \cdot M - \tau_2) - \tau 1)\end{aligned}$$

where $W_1, W_2, V_1, V_2, U_1, U_2$ are matrices with coefficients drawn randomly from a uniform distribution between -1 and 1, as are also the vectors $\mu_1, \mu_2, \nu_1, \nu_2, \tau_1, \tau_2$. This is equivalent to the choice of the measurement unit for the signals of our model. σ is an arbitrary nonlinearity, here the hyperbolic tangent function. The $C_{i,k}$ are drawn from a centered normal distribution whose variance, which can be understood as the size of the retina, was so that the sensory changes resulting from a rotation of the eye were of the same order of magnitude as the ones resulting from a translation of the eye.

$Q = (Q_1, \ldots, Q_q)$	positions of the joints
$P = (P_1, \ldots, P_p)$	positions of the eyes
$a^\theta_i, a^\varphi_i, a^\psi_i$	euler angles for the orientation of eye i
$Rot(a^\theta_i, a^\varphi_i, a^\psi_i)$	rotation matrix for eye i
$C_{i,k}$	relative position of photosensor k within eye i
$d = (d_1, \ldots, d_q)$	apertures of diaphragms
$L = (L_1, \ldots, L_p)$	positions of the lights
$\theta = (\theta_1, \ldots, \theta_p)$	luminances of the lights
$S^e_{i,k}$	sensory input from exteroceptive sensor k of eye i
S^p_i	sensory input from proprioceptive sensor i
M, E	motor command and environmental control vector

In Organism 1, motor command M is a vector of size 40, and S a vector of size 40. The organism has 4 joints and two eyes with 20 photosensors each. The eyes are free to rotate in any way, including through torsion movements. θ and d are constants drawn at random in the interval $[0.5, 1]$.

Figures 3.b and 3.c present the results obtained for a linear approximation in a neighborhood of size 10^{-8} (this is to be understood with respect to the unit measure given by the previous matrix coefficients). Sensory inputs were generated from 50 motor commands and 50 environmental positions following a normal distribution with mean zero and standard deviation 10^{-8}. Coordinates differing from zero by more than the standard deviation were put equal to zero. This neighborhood size is the one that yielded the most significant separation between very small eigenvalues and non-null ones, but results could be found from about 10^{-5}. Generally speaking, the size of validity for the linear approximation is of course related to the curvature of the sensory manifold and there is no general answer to the question of what size the neighborhood should have.

Out of these 50 motor commands $\{dM_i\}$ and 50 environmental positions $\{dE_j\}$, we produced 50 sensory inputs changes $\{dS_i\}_{dE=0}$ resulting from the world configurations $\{(M_0 + dM_i, E_0)\}$, 50 sensory input changes $\{dS_j\}_{dM=0}$ resulting from the world configurations $\{(M_0, E_0 + dE_j)\}$, and $50 \times 50 = 2500$ sensory input changes $\{dS_{i,j}\}$ resulting from the world configurations $\{(M_0 + dM_i, E_0 + dE_j)\}$.

The results of the other two experiments (Figure 4) were obtained with environmental lights having variable luminance. The organism had 10 joints, and a diaphragm for each of its 4 eyes. Motor commands were of dimension 100. For Organism 2 the diaphragm aperture d_i of eye i was defined by the equation:

$$\sum_k S_{k,i} = 1$$

that is, the aperture adjusted *automatically* so as to ensure constant total illumination for eye i. For Organism 3, d was under its own control:

$$(Q, P, a, d) = \sigma(W_1 \cdot \sigma(W_2 \cdot M - \mu_2) - \mu_1)$$

Theoretically the eigenvalues of the covariance matrices of the three sample sets $\{dS_i\}_{dE=0}$, $\{dS_j\}_{dM=0}$, and $\{dS_{i,j}\}$, should fall into two classes, one with zero values, and one with non-zero values. To distinguish these two classes we used a clustering method, and supposed that the two classes V_1 and V_2 were such that each λ of V_1 was more comparable in size to other λ's of V_1 than to all those of V_2, and conversely. Finding the boundary between the two classes can thus be done by ordering the λ_i in decreasing order, and locating the value of i such that the ratio between λ_i and λ_{i+1} is largest (See Figures 3.c and 4). We could also have used an approach similar to Minka, 2000. It should be noticed that the nullity or

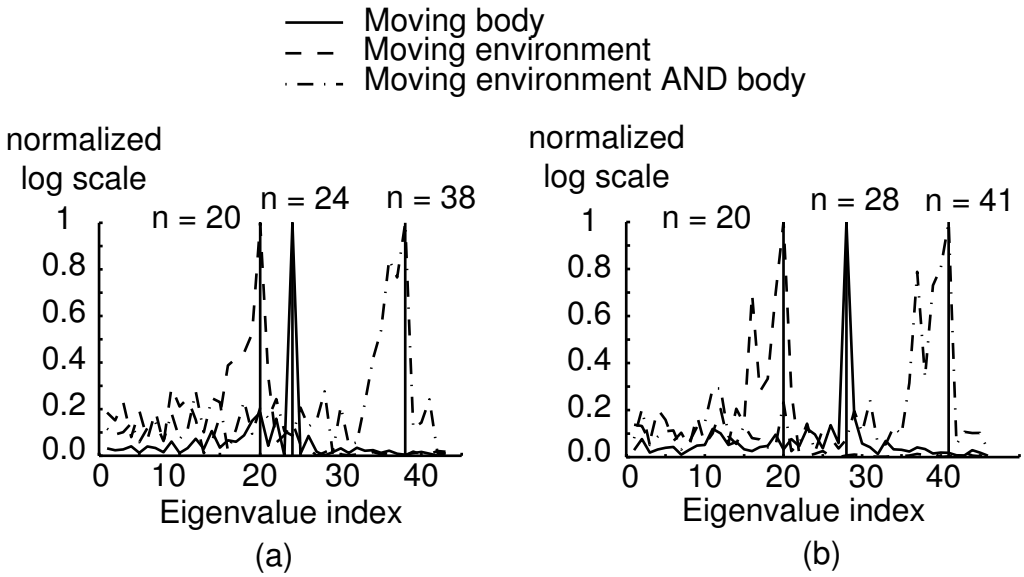

Figure 4: Dimension estimation when the diaphragm aperture is determined by reflex (a) and controlled by the organism (b).

non-nullity of the eigenvalues reflects a characteristic of space, but their absolute value reflects a characteristic of the sensorimotor system. This is where lies the practical aspect of our problem as compared with the theoretical one.

7.2 Extensions of the model

The concrete examples we have presented here might give the impression that the approach is limited to cases where motor commands and sensory input are defined as instantaneous vectors — this would be unrealistic, since motor commands and neural afference are usually extended in time. But this poses no problem for our approach, since $\mathcal{M}$, $\mathcal{E}$ and $\mathcal{S}$ need only to have a manifold structure, not a vectorial structure. For simplicity we chose vectors representing the motor commands and sensory inputs at a particular time t, but in general M and S could be vector functions, and the sensorimotor relation would then be an equation of functionals. In the finite case, our approach could be applied directly, and performing the PCA would then require making a local approximation of this functional relation in a manner analogous to that proposed by Matarié & Jenkins, 2000 for motor control. In the non-finite case, we think there should also be a way to use our method,

since $\mathcal{C}(M_0, E_0)$ is finite even if $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{E}$ are not.

When the functional equation is given as a differential equation, the sensorimotor relation remains instantaneous. We will in general be able to write it in implicit form:

$$\psi(S(t), \dot{S}(t), \ldots, M(t), \dot{M}(t), \ldots, E(t)) = 0 \tag{5}$$

and then determine the variables that are functions[5] of $S(t), \dot{S}(t), \ldots$ which can be accounted for in terms of $M(t), \dot{M}(t), \ldots$. From there on we can apply the same reasoning we used for the exteroceptive variables described in the present paper. This extension is one of the objects of our current research.

[5]Linear combinations, in the case of a local linearization, which could, again, be determined by a PCA. Such a linearization could perhaps also be applied globally, since linear differential equations can be used to explain a wide range of physical problems, and in particular probably those that can be apprehended quickly by our brains.