Perception as a dynamical sensori-motor attraction basin

M. Maillard[1], O. Gapenne[2], L. Hafemeister[1], and P. Gaussier[1]

[1] Neuro-cybernetic team, Image and signal processing Lab., ETIS UMR CNRS 8051
Cergy Pontoise University / ENSEA, 6 av. du Ponceau, 95014 Cergy , France email:
`maillard@ensea.fr`, Home page: `http://www.etis.ensea.fr/~neurocyber`
[2] COSTECH Université de Techonologies de Compiègne groupe suppléances
perceptives email: `olivier.gapenne@utc.fr`, Home Page:
`http://www.utc.fr/gsp/accueil.html`

Abstract. In this paper, we propose a formal definition of the perception as a behavioral dynamical attraction basin. The perception is built from the integration of the sensori-motor flow. Psychological considerations and robotic experiments on an embodied "intelligent" system are provided to show how this definition can satisfy both psychologist and robotician point of view.

1 Introduction

The classical conception of perception refers to a parallel and passive computation of an input flow of information. In this frame a cognitive system is considered only as a computational system receiving inputs (namely the "sensations") to identify objects or events and producing output representations useful for reasoning and leading to appropriate actions. In this paper we investigate an alternative approach where perception is linked to dynamical laws between actions and sensations [22, 14]. But the lack of a formalism leaves the community without tools to analyse this kind of concept. Therefore, taking avantage of a generic formalism developped to analyse cognitive systems [5], and refeering to psychological and neurobiological assumptions, we propose to define the perception as a dynamical sensori-motor attraction basin. Specifically, the formal description of a sensori-motor system leads us to define the perception in regard of the dynamics of the sensori-motor system. Moreover, we will show that a sensori-motor learning in a competitive neural network allows to approximate such an attraction basin. The immediate benefit is to be able to explore the concept by observing the resulting perception in robotic experiments and to confront it to psychological experiments.

2 Background considerations

We briefly wish to recall that our formalism of perception inherits from a recent philosophical and scientific tradition. Indeed, during the 19th century, and more definitively during the 20th century, a whole philosophical work [9, 17] states that the conscious experience and knowledge are the fact of a construction (third person) or constitution (first person). This poses that to know and perceive in an organized way is not given and supposes developmental and learning processes. Moreover, this epistemology of the construction or constitution which renews the statute of objectivity deeply affirms the role of the action and its effects as a condition of possibility and constraint [16]. And very radically, it poses that the lived experience of the world and oneself (and their relation) are defined by the properties of the actions system available to the agent which organizes the organism-environment relations. This led to the idea that the causality of the experience cannot be reduced to a strictly active or passive internal construction. In the scientific field, this approach had some eminent representatives so much in the life sciences [12, 4] that in the social sciences [15, 21, 13]. There is obviously no question of developing the whole of this work here. We simply propose to specify the importance of the action and more precisely of some to know about this same action so that the process of perceptive genesis can take place. The perceptual knowledge associated with the action is classically described like concerning the proprioception. The latter indicates a system of coupling which intervenes in the perception of the movement (kinesthesis) and body positioning (statesthesis). Like any system of coupling, it implies a whole of elements which are not limited to particular sensors (neuro-muscular spindles, neuro-tendinous bodies or articular sensors) and their mode of transduction. This system implies a nervous network (cortex included in particular sensory-motor, premotor, left parietal, cingular bilateral cortex and supplementary motor area), effector (the muscles) and includes the environmental constraints (gravity, friction). Without going into the details of the operation of this coupling, it is possible to mention some significant points:

– Each movement is associated to a specific reafferent sensory flow which can be defined like a true signature.
– This signature is currently described in the shape of a vector which includes acceleration, speed, direction and duration of each movement.
– Formally, the relation between movement and sensory reafferences is bijective what guarantees a great stability of the invariants that this coupling authorizes.
– This system is continuously activated but some forms of adaptation at the time of prolonged immobility can be observed.

The proprioceptive coupling thus allows the constitution of reliable invariants relative to the body by convening the body itself. This sensitivity of the proprioceptive system to the only directed deformations of the body confers a particular statute to him to the glance of two other systems (the vestibular and tactile systems) implied in the general sensitivity to the movement and the position. These two systems can indeed generate flows independently of the active or passive deformation of the body. Being given this specific property, it is not surprising that Roll [19] suggested the founder role of this system in the emergence of one experienced body. In addition to the very early functionning of the motor-proprioceptive loop during the foetal life, the question of the basic statute of the proprioception does not seem to be any doubt if one considers work of Held and Hein [8] and especially of Buisseret and collaborators [2]. However, and very basically, it is significant to consider that, on the level more strictly empirical, the description of the constitutive role of the action remains problematic. Indeed, the control condition (absence of action) can be only exceptionally satisfied and with much difficulties what justifies besides the possible recourse to artifacts like robots to validate radically constructivist assumptions. Moreover, it should be considered that this integrating statute of the proprioception is frequently threatened by reafferent co-occurrent flows which can introduce confusions (exteroceptive flow associated the displacement of the environment) or supplant this statute (exteroceptive flow associated to the movement of the body itself which can induce vection illusion). Thus, to pose the funding statute of the proprioception seems admissible only in the terms of a developmental process. We have to note that these important questions are not directly studied in robotic research. Moreover, the movement situates the subject in a temporal unit which resounds on a multitude of natural coupling systems. The unicity of the action is a vector of multimodal integration by way of redundancy, of complementarity. And this point echoes two concepts, suggested by Gibson [7], complementary and extremely relevant to advance in the way of a formalisation of the perceptual learning. These two concepts are those of proprioceptive function and co-perception. Briefly, the first one suggests that the whole of the sensory flows (visual, tactile, vestibular, auditive, etc...) associated to the movement intervene at the same time in the regulation of postural tonicity and the experienced body one. The second one is further posing that, if these flows intervene in the constitution of one oneself (ecological self), they specify simultaneously the world. The proprioceptive function mainly implies low spatial and high temporal resolution (peripheral visual way, tactile spinothalamic way, etc...) sensors of flow. A flow can be defined like the continuous variation of a source of energy on the surface of a sensor. The variation is necessary and it is related to the variation of the sensor position and orientation and/or to the variation of the afferent sensorial flow. The relevant point for the subject is to be able to dissociate these sources of variations; the question of the agentivity defined as the capacity of an agent to perceive that some transformations of the world are directly tied to its proper action is posed at this level. In fact, the organism has signals permanently relating to its positioning in space (deep muscular sensitivity, angular values of the articulations). Exteroceptive flows will be associated, integrated into this major sensitivity. This is the coordination of these two flows which constitutes at the same time the proprioceptive function [3] and co-perception [18]. And the possible detection of temporal coincidences between these two flows constitute the base of the learning of the regularities within the sensory-motor loops (importance of the spatio-temporal redundancies). The formalism that we present at the continuation is inspired very directly by these conceptual evolutions specific to a better understanding of the developmental processes which link motor-sensorial loop, proprioceptive function and perception.

3 Mathematical definition of Perception

Previous works have focused on mathematical tools to formalize pure behaviorist or reactive systems (see Steels, Smithers, Wiener, Ashby, etc...). The most interesting tools are presented in the classical NN literature and in the dynamical system theory[20]. But these tools are dedicated to specific parts of what we will call a cognitive system (CS)[3]. We summarize here the basis of our mathematical formalism of CS. A CS is supposed to be made of several nodes or boxes associated with input information, intermediate processes and outputs (command of actions). Each element presents a degree of complexity that ranges from a simple scalar product (or distance measure) to a more complex operator such as an "If...then...else..." statement (hard decision making), a recurrent feedback in the case of a dynamical system, a mechanism to control focus of the attention, etc... Whatever the complexity of an element is, we state it as a "neuron". The inputs and outputs of a CS are represented by vectors [4] in the "bracket" notation[5]. An input or output vector noted $|x\rangle$ with $|x\rangle \in R^{+m}$ while its transposed vector is noted $\langle x|$. Hence $\langle x|x\rangle$ represent the square of the $|x\rangle$ norm. We can consider that any element of a CS filters an input vector according to a matrix of weights A and a non linear operator $\mathbf{k}$. This operator represents the way to use the A matrix and the pattern of interactions between the elements of the same boxe. For instance, in the case of a simple WTA (Winner Takes All) boxe, its output $|y\rangle$ is $wta(A|x\rangle)$ with $|y\rangle = (0, ..., y_j, ...0)$ and $j = ArgMax(q_i)$ and $q_i = \langle A_i|x\rangle$. Different kinds of inputs/outputs connections with their weight summarized in the matrix A exist. Basically, we distinguish 2 main kinds of inputs/outputs connections:

- "one to one" connections named I used to transmit information (unconditional stimulus US) from one group to another one, and seen as a reflex pathway which can not be affected by learning.
- "one to all" connections used for transmiting conditional stimuli (CS), having learning capabilities and used for categorization, etc...

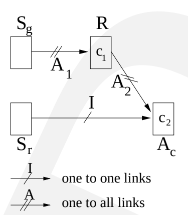
Finally in the case of a complex competitive and conditioning structure with 1 unconditional (US) and 2 conditional (CS) inputs, we simply write $|y\rangle = c(A_1|CS_1\rangle, A_2|CS_2\rangle, I|US\rangle)$. This allows to be sure a particular matrix is always associated to the correct input vector but it does not mean the matrix has to be multiplied by the vector (this computation choice is defined by the operator

[3] The term cognitive is used in the sense of the study of particular cognitive capabilities (cogitare - to think) and does not induce any a priori cognitivist approach.

[4] We consider the components of the different input/output vectors can only be positive/activated or null/inactivated. Negative activities are banned to avoid positive effects when combined with a negative weight matrix.

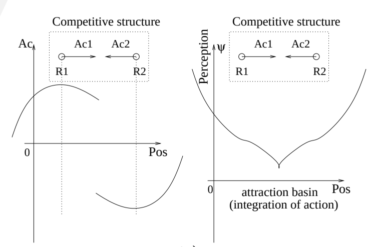
[5] The choice of this notation will be explained in the conclusion of the paper

itself). Interestingly this formalism emphasises the fact that an operator is working on 2 different flows of information moving in opposite directions. The first one transforms sensory information in an output code while the second one acts on the group memory in order to maintain an equilibrium defined by the learning rule. Reaching an equilibrium allows the system to have a stable behavior according to its environment but also to adapt itself to environmental variations. These properties echo the psychologic assumptions previously overviewed and show that a percept can only be built if there are interactions between a subject and his environment. Based on these considerations we propose to define the perception as a dynamical attraction basin allowing stable behavior through time. In robotic homing experiments, a similar phenomenon was already observed as learning the construction of a behavioral attraction basin surrounding the goal is enough to allow the robot to return to a place without being able to statically recognise it [6]. Also in visual perception [7], an affordance can be defined as building or accessing to an invariant caracterizing one particuliar sensori-motor behavior. In this case, perception is considered as the result of the learning of sensations/actions associations allowing a globally consistent behavior while facing an object.
For instance, let us consider a sensori-motor system of an agent acting in a given environment (or state space), and having 2 sensation vectors $|S_r\rangle$ and $|S_g\rangle$. First, $|S_r\rangle$ represents the proprioception, a coarse feedback information from the execution of the motor command or the direction of the goal (if the goal is in the immediate neighborhood). It can be considered as a reflex or regulatory pathway linking a proprioceptive sensor to the motor command $|Ac\rangle$. Second, $|S_g\rangle$ represents a more global information about the environment allowing to build a local but robust distance measure (metric). This measure is learned and computed by a competitive recognition group R ($|R\rangle$ representing its output activity). This basic sensori-motor architecture, used as well for a homing task as for a focus on a target, can be described by the diagram fig. 1a and its corresponding equations. The operator c represents a competitive structure (soft-WTA) able to



$$|R\rangle = c_1(A_1|S_g\rangle)$$
$$|Ac\rangle = c_2(A_2|R\rangle, I|S_r\rangle)$$

a)



b)

Fig. 1. a) An exemple of a simple sensori-motor system and its corresponding equations. b)*left* Theoretical system actions after learning 2 sensation/action associations and their competition according to the system spatial position. b)*right* Theoretical perception and attraction basin computed by integration of the action shown on left.

self-organize itself according to one sensory data flow (also in the case of c_2 fig. 1, to condition one input data flow according to an unconditional flow). After the competition, the activity of R reflects the recognition level. The decision is delayed at the motor level and must be understood according to the global temporal dynamic of the system. Let us notice the system behavior does not directly depend on the absolute level of recognition of the learned views or places. Only the rank in the competition process matters which allows some robustness to perturbations until the noise as an effect on the rank in the competition. The output of such a sensori-motor system is an action realized by the agent in its environment. Consequently, the agent modifies its state and then its sensations: the agent and its environment can be viewed as a dynamical system [1]. It is important to notice that in the dynamical systems theory, the action is defined as the derivative of a potential function [10, 20]. Considering the psychological background (section 2), we can precise the definition of the perception we gave as an attraction basin: we call this potential function *perception*. Consequently, agent's actions derivate from its state of perception. Formally action $|Ac\rangle$ can be deduced from perception Per with this relation :

$$|Ac\rangle = -m\ \overrightarrow{grad}\ Per(\boldsymbol{p}) \tag{1}$$

where $\boldsymbol{p}$ is the agent state and m is associated to a "virtual" mass which deals with the agent's embodiment. For instance, m should change while considering two different agents with different morphologies. This mass which allows homogeneity of equation(1) will be considered as a constant in the following. Of course equation (1) can be rewritten: $Per(\boldsymbol{p}) = -\frac{1}{m}\langle Ac|\boldsymbol{p}\rangle = -\frac{1}{m}\int_{\boldsymbol{p}+\delta\boldsymbol{p}} Ac\ d\boldsymbol{q}$. which corresponds to our intuition of the recognition as an attraction basin. An illustration of a perception basin can be viewed on fig. 1b-*right* where only two actions were possible ("go left" and "go right"). The basin results from the numerical integration of the curve proposed fig. 1b-*left* and represents the learned sensori-motor associations and their effects. According to the agent point of view, its perception is built from the integration of its actions relative to its state $\boldsymbol{p}$. In consequence $\boldsymbol{p}$ refers to the agent proprioception and other internal variables which can be implicit in the system. But considering an external point of view, the integration of the agent's actions relative to his spatial position allows to visualize a posteriori a perception which keeps meaning: we obtain a visualization of the dynamical behavioral attraction basin of the agent.

4 Robotic Application

In order to appropriate our definition of perception and understand its full details, we propose a simple robotic application using a Koala robot with a CCD camera. The robot task is to learn how to return to a given object which can be interpreted as the fact the robot "perceives" the object. The robot only learns affordances [7] linked to the target; an explicit recognition of the object is not required. During the experiment, we propose to evaluate the global behavior of the robot by observing its trajectories. An external observer we will conclude that our robot succeeds in its task if its global behavior is consistent while facing the learned object: a particular affordance is then observed. We also propose to a posteriori compute the robot's perception according to the equation (1). These two observations are in fact complementary: the first one focus on a robot trajectory, which corresponds in a way to descend the perception basin computed by the equation (1).

In order to provide useful and simple sensori-motor associations, the visual features extracted from the visual flow require some robustness as a function of the robot task. As the robot has to move towards an object in unknown environmental conditions, it has to face large non linear transformations of the images (scale, perspectives, etc...). To achieve scale, contrast and luminance invariance, keypoints are extracted on the input images by a multi-scale algorithm inspired by Lowe's work [11]. They fit with the local extrema of the scale-space images filtered by DOGs (Difference Of Gaussians). Finally, at each keypoint, a local feature is coarsly extracted: the first two moments of the orientations of the gradient image in the 4 neighbourhoods of each keypoint are kept. These informations are gathered from all the scales and normalized on neuronal maps which constitute the visual sensory data of the sensori-motor system.

In order to learn object affordances, the robot associates its actions to reach the target object with its sensori data (here visual information). This learning phase is supervised as the direction of the action is provided by an operator via a joystick (fig. 2). This unconditional stimulus corresponds to the "proprioception" pathway reported in section 2. Finally, the association is performed by a conditioning mechanism based on the classical LMS (Least Mean Square algorithm) [23]. The decision about the action is given by a competitive neural group (WTA). We can notice that the capacities of (linear) separation of such a mechanism are coarse. A simple way to generalize this mechanism in a task involving multiple objects, is to duplicate the group called LMS (for instance one per object selected by the context given by the operator even if it is not a good solution in term of efficiency and biological plausibility). We verify that only a few sensori-motor associations enable the robot to reach the learned object.

As proposed previously, fig. 2 shows one of the robot trajectory recorded after the learning phase was acheived. We can notice that the robot has a direct trajectory when the learned object is in its visual field but zigs zags when the object is not. As the robot doesn't have an explicit object recognition module nor a tracking one, it can't search for the target absent of its visual field. When this happens it just makes its way toward what looks similar to the learned object relative to its perception. Due to the robot movements, the target eventually can be seen again as in fig. 2. Of course the robot visual field is modified by the dynamic of the sensation/action/environment loop.

Finally, we propose to compute the a posteriori perception of the robot. The state of the robot is defined by its spatial location in the environment and by its body and CCD camera orientation relative to the learned object. In order to record the action at each spatial location, we impose the robot's trajectories. Then according to equation (1), perception is given by the integral of the robot

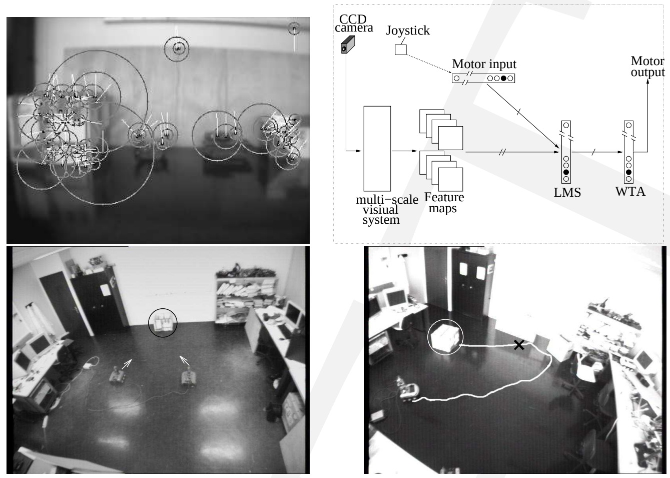
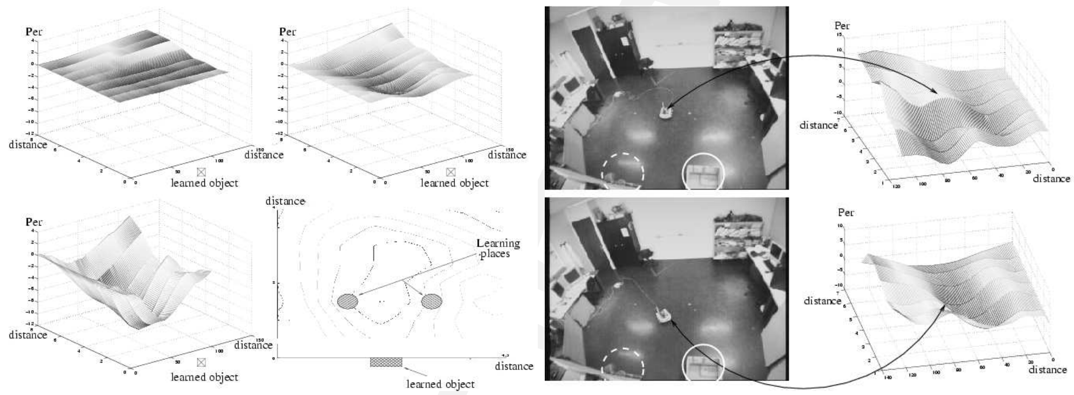


Fig. 2. *Top Left* : Example of an input image and of its keypoints (the circle size indicates the scale where the keypoints were found) *Top Right* : Overview of the architecture with the conditioning mechanism. *Down Left* : Example of a learning phase: the circled object has to be learned. *Down Right* : A robot trajectory (white line). The black cross represents the place where the target (the circled box) enters in the robot's visual field.

action over space. Several measures with a constant orientation of the CCD camera have been successively made during a learning phase. The learning can be considered like the creation of a dynamical attraction basin. The latest is created only after few sensori-motor associations. The shape of the final basin (fig. 3-a) fully explains the globally consistent behavior of the robot: the robot "falls" down the perceptive basin and consequently reaches the learning object. These measures corroborate in a quantitative way the results observed on trajectories. Furthermore the good generalization capability of the given architecture can be easily displayed out of the learning places.

The sensory data extracted are coarse and their small number eases the learning of the sensori-motor associations without allowing a good discrimination of objects. While facing the object which was learned previously and an object which was not, neurons activities coding the robot's action are quite similar. Even if the learning allows to dig a deeper basin in the case of the learned object, the difference in depth between the attraction basins we compute is too small to explain the global consistent behavior of the robot: it reaches its learned target independently of its starting spatial location (assuming the learned object is in its visual field). In fact the previous measures do not take into account all the dynamic aspect of the system. In particular, the robot modifies its orientation (and so its sensations) relative to the object according to its state of perception while moving towards the object. These movements allow to disambiguate the visual flow (figure 3-b). This dynamic of the sensori-motor loop ensures a con-



(a) (b)

Fig. 3. *(a)* : Visualization of the perception at different learning steps. The different curves of iso-perceptions (right down) associated with the drawing of learning places underline the system capacity to generalize. *(b)* : Visualizations of the perception depending on the orientation of the robot. Top : 2 basins are present even if just the object box (full-line circled) was learned. The object lamp (dot-line circled) creates the second basin. Down: the robot's orientation is modified after a few of its actions. The robot is in a new state of perception without ambiguity

sistent behavior which is impossible to obtain in a static way. This experiment clearly shows how essential it is to consider the 3 dimensions of the problem (the spatial location (2 dimensions) and the orientation relative to the object). Only then the notion of a sensori-motor invariant can be grasped but unfortunately the 4D basin cannot be drawn.

5 Conclusion

This paper is an attempt to fill the gap between the psychological concept of perception and the dynamic of a sensori-motor system and the behavior of a robot acting in its environment by proposing a formal definition and an experimental measure of the perception. In this context, learning some particular affordances can be seen as building an attraction basin. A system is in a *stable state of perception* if it is able to maintain itself in the associated attraction basin. Hence recognizing an object (from visual, tactile, auditory... informations) is seen as choosing to stay in a given basin. The choice of the formalism inspired by quantum mechanics (vectors noted with brakets) is linked to the idea that manipulating vectoriel information represents somehow a wave function. An observation of the perception by the agent could be seen as to trigger one particular behavior or in other words to freeze a state of perception (attraction basin). Taking into account this dynamic extends the frame of active vision since the agent becomes an active actor itself. Therefore future works will study how could the agent autonomously construct its own perception of its envronment.

References

1. R.D. Beer. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4):209–243, 2003.
2. P. Buisseret, E. Gary-Bobo, and P. Imbert. Ocular motility and recovery of orientational properties of visual cortical neurons in dark-reared kittens. *Nature*, (272):816–817, 1978.
3. A. Bullinger. *Du nouveau né au nourisson*, chapter Vision, posture et mouvement chez le bébé: approche développementale et clinique, pages 48–61. F. Jouen and A. Henocq, Paris, 1991.
4. A.K. Engel and P. König. *Philosophy and the cognitive sciences*, chapter Paradigm shifts in neurobiology: towards a new theory of perception, pages 131–138. Wittgenstein-Gesellschaft, Kirchberg, 1993.
5. P. Gaussier. Toward a cognitive system algebra: A perception/action perspective. In *European Workshop on Learning Robots (EWLR)*, pages 88–100, 2001.
6. P. Gaussier, C. Joulain, J.P. Banquet, S. Leprtre, and A. Revel. The visual homing problem: an example of robotics/biology cross fertilization. *Robotics and Autonomous Systems*, 30:155–180, 2000.
7. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, 1979.
8. R. Held and A. Hein. Movement produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, (56):872–876, 1963.
9. E. Husserl. *Things and space: Lectures of 1907*, volume 7. Springer, 1998. english translation of Dingvorlesung.
10. J.A. S. Kelso. *Dynamic patterns: the self-organization of brain and behavior.* MIT Press, 1995.
11. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, pages 91–110, 2004.
12. H. Maturana and F.J. Varela. *Autopoiesis and cognition:The realization of the living.* Reidel, 1980.
13. J.K. O'Regan and A. Noe. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, (24):939–1031, 2001.
14. R. Pfeifer and C. Scheier. Sensory-motor coordination: the metaphor and beyond. *Robotics and Autonomous Systems*, 1996.
15. J. Piaget. *La naissance de l'intelligence chez l'enfant.* Delachaux et Niestle Editions, Geneve, 1936.
16. H. Poincaré. *Science and Hypothesis.* Walter Scott, 1905.
17. M. Merleau Ponty. *Phénoménologie de la perception.* Gallimard, 1945.
18. P. Rochat. The self in early infancy. *Theory and research, advances in Psychology*, (112), 1995. Elsevier Science Publishers.
19. J.P. Roll. Physiologie de la kinesthese. la proprioception musculaire: sixieme sens ou sens premier. *Intellectica*, pages 49–66, 2003.
20. G. Schöner, M. Dose, and C. Engels. Dynamics of behavior: theory and applications for autonomous robot architectures. *Robotics and Autonomous System*, 16(2-4):213–245, December 1995.
21. E. Thelen and L.B. Smith. *A dynamic system approach to the development of cognition and action.* MIT Press, 1994.
22. F. Varela, E. Thompson, and E. Rosch. *The Embodied Mind.* MIT press, 1993.
23. B. Widrow and M.E. Hoff. *Adaptive swithing circuits.* IRE WESCON convention records, 1960.