Categorization in a real-world agent using haptic exploration and active perception*

Christian Scheier and Dimitrios Lambrinos

AI Lab, Computer Science Department, University of Zurich
Winterthurerstrasse 190, 8057 Zurich, Switzerland
E-mail: {scheier,lambri}@ifi.unizh.ch

Abstract

An agent in the real world has to be able to make distinctions between different types of objects, i.e. it must have the competence of categorization. In mobile agents categorization is hard to achieve because there is a large variation in proximal sensory stimulation originating from the same object. In this paper we extend previous work on adaptive categorization in autonomous agents. The main idea of our approach is to include the agent's own actions into the classification process. In the experiments presented in this paper an agent equipped with an active vision and an arm-gripper system has to collect certain types of objects. The agent learns about the objects by actively exploring them. This exploration results in visual and haptic information that is used for learning. In essence, the categorization comes about via evolving reentrant connections between the haptic and the visual system. Results on the behavioral performance as well as the underlying internal dynamics are presented.

1 Introduction

The ability to make distinctions is fundamental to adaptive behavior. It is one of the basic competences of any agent in the real world. Animals need to discriminate between food and non-food, they must recognize the nest, and they have to differentiate between conspecifics and individuals of other species. A robot for collecting garbage has to be able to distinguish between garbage and non-garbage, the garbage truck from other trucks, other garbage robots from human agents, it has to be able to recognize the charging station, etc. This competence is called classification or categorization. Some aspects of this capacity can be innate or pre-defined such as the taste of particular foods or the recognition of conspecifics. Others, like the shape of the food items and their location, might vary greatly, depending on the kind of environment. While a garbage robot might have a predefined way of identifying the charging station or of recognizing the garbage it should collect, the location of the charging station, or the visual appearance of the garbage might vary from city to city. There are several ways to endow a robot with the capacity to categorize. One way is to predefine the categories by either incorporating a special sensor (e.g. for color) or by predefining configurations in the sensory space that correspond to various categories. While the former is easily doable, we do not gain many insights into the nature of intelligence. The latter turns out to be difficult, especially if we take into account that the agents move around and that one and the same object leads to a large variety of proximal stimulation of the sensors. As an example, consider the visual space of the robot that was used to conduct the experiments presented in this paper. It is equipped with a CCD camera with approximately 2^{18} photoreceptors where each photoreceptor has 2^8 states, i.e. there are $2^{2^{21}}$ different states . Clearly, the problem of reducing the many degrees of freedom of this space is a non-trivial one. It is extremely difficult to define meaningful a priori configurations within this space. This implies that categories should be acquired at run time, i.e. as the agent is moving around in the environment.

Traditionally, categorization has been treated as an information processing question: the sensors receive a particular input which is processed and mapped onto an internal representation (e.g. a category node). This view dominates most psychological models of categorization. In these models categorization usually consists of associating some binary input vector to a category node via supervised learning schemes (see e.g.[6]). Similarly, computer vision systems generally have considered the problem as a passive or static one where there is not definite relation between the object and its environment (e.g. [13]). In these systems categorization is seen as a problem of matching the visual input to a stored rep-

*AILab, Tech. Report No. 95.15, submitted to SAB96 (FROM ANIMALS TO ANIMATS, Fourth International Conference on Simulation of Adaptive Behavior, Cape Cod, Massachusetts, USA, September 9-13, 1996)

resentation or model of objects using methods such as direct template matching, hierarchical template matching or transform and match (see e.g. [5] for an overview). The problem encountered is that typically an enormous number of highly different sensory patterns should map onto the same representation. Efforts to solve this problem have only been successful to a limited extend. Biological systems, on the other hand, are extremly efficient at categorizing their environment at run time (see e.g. [4]). One key difference between information processing approaches and categorization in natural systems seems to be that the former view categorization as a process which happens on the (visual) input side only whereas the latter heavily includes information stemming from the organism's own movements. For recognizing a memorized visual stimulus, for instance, flies have to shift the actual image to the same location in the visual field where the image had been presented during the storage process ([2]). In other words, the incorporation of action is crucial for the recognition process.

Self-motion is even more important for the development of categories in young infants. In the latter half of the first year, infants begin to actively manipulate the objects that they explore visually using several explorative activities such as mouthing, fingering, rotating or banging behaviors (e.g. [9]). This manipulation has the important consequence that the infant may acquire tactual and kinesthetic (i.e. haptic) information about the object through active touch. Recent studies indicate that during the first semester, infants can discriminate haptically among objects and transfer information from one modality to another (see e.g. [9] for an overview). Thus, from the earliest age, hands and mouth can be used to pick up information about objects. In coordination with other perceptual systems, the hands and mouth contribute to the perception of objects in the environment. There are active cross-modal comparisons, i.e. infants actively relate what they experience of the object in the haptic and visual modalities ([12]). In sum, infants discover object categories through the cross-correlation of multimodal experiences. Moreover, they disover such categories by acting upon the objects and exploiting the resulting multimodal experiences of objects (i.e. exploration). In our previous work we have developed an approach to adaptive categorization in mobile robots that is based on these results ([7],[8],[10],[11]). The main idea is to view categorization as a *sensory-motor coordination* rather than an isolated perceptual (sub-)system. This is achieved by including the robot's own actions into the classification process. In these experiments we have demonstrated how these ideas can be used to enable an agent to reliably learn a classification of objects in a straightforward and simple way. In this paper we considerably extend this framework. First, the sensory-motor complexity of the robot is significantly increased.

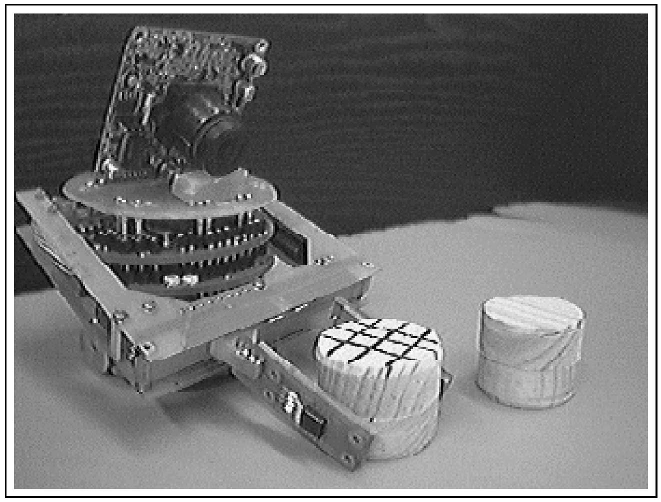

Figure 1: The robot and its environment. Explanations see text.

While the sensory-motor system in the previous experiments consisted of IR sensors and two wheels the robot used in this paper is additionally equipped with a CCD camera and an arm with a gripper mounted at the end. Instead of passively scanning the camera an artificial eye has been implemented which actively moves a fovea to interesting parts (e.g. bright spots, texture, movement) of the image. Once it has visually focused an object the robot starts exploring it using the arm-gripper system. Similar to what has been observed in infants, the robot uses the resulting multimodal (i.e. haptic and visual) experiences to learn about important properties (e.g. texture, conductivity) of the explored object.

A second improvement concerns the categorization mechanisms used. Previously, categorization was based on (a) learning a sensory-motor mapping using a temporal Kohonen map [7] or growing dynamical cell structures ([8]) and (b) associating this mapping with behavioral processes such as grasping, pushing or avoiding. The basic categorization mechanism was a conditioned association between the learned sensory-motor mappings and some behaviors. In the experiments presented in this paper categorization is not achieved via conditioning but rather by a learned *reentrant mapping* between visual and haptic feature maps. The term *reentry* refers to the fact that there are reciprocal connections between the feature maps. Reentry is necessary to account for the coordination of responses across modalities. When the robot explores an object the experience is visual, but also invokes haptic stimulation such as hardness or roughness of the object. Reentry is a mechanism to correlate these perceptual informations on the basis of their temporal contiguity. Thus, categories develop from the real-time correlations that exist accross the independent stimuli. This *correlative* function of reentry has been suggested to be one of the core mechanisms for the de-

velopment of categories in infants ([12], see also [3]). A final extension of our previous work is the concept of *attentional sensory-motor loops* which are modulated by the category-specific responses of the reentrantly connected feature maps. In essense, categorization consists of breaking or enhancing the attentional sensory-motor loops depending on the type of object encountered and the resulting activity in the feature maps. Each of these mechanisms will be explained in more detail below. The *task* of the robot is to collect certain types of objects. Since there are other objects in the environment which it should not collect, it has to learn (a) to recognize objects it should collect and (b) to discriminate the latter from all other objects in the environment.
The paper is structured as follows. First, the experimental set-up is described. Second, an overview of the architecture as well as a detailed description of the visual and the haptic system are given. Third, the results of several experiments are presented. Finally, the main problems and future work are pointed out.

2 Experimental Set-up

The mobile robot used in the experiments is is a $Khepera^{TM}$, 55mm in diameter and 32mm high (weight 70g) (see figure 1). The effector system consists of two wheels which are individually driven by DC motors, and an arm with a gripper installed at the end. The maximum object size that can be grasped with the gripper is about 40mm. There are two types of objects in the environment (see figures 1 and 2). Wooden objects with texture on the surface and others without texture. Textured objects have been made conductive by rapping a metalic wire around them. The sensory system consists of a visual and a haptic system.

2.1 Visual system

Input to the visual system is provided by a miniature (1/3", 18g) B/W CCD camera (ces VPC-465). The camera is equipped with a built-in lense with 90 degrees viewing angle. In order to reduce computational load all visual processing was done on a visual server based on a Pentium 133Mhz PC. The video image is sampled at a maximum of 30 fps using a video framegrabber (PMS from Media Vision). Images were grabbed at resolution of 640 x 480 pixels and were spatially averaged to provide an image of 160 x 120 pixels (see figure 2). The main part of the control architecture was run on a workstation (SUN SPARC 10). Communication between the visual server and the workstation was done using the TCP/IP communication protocol.

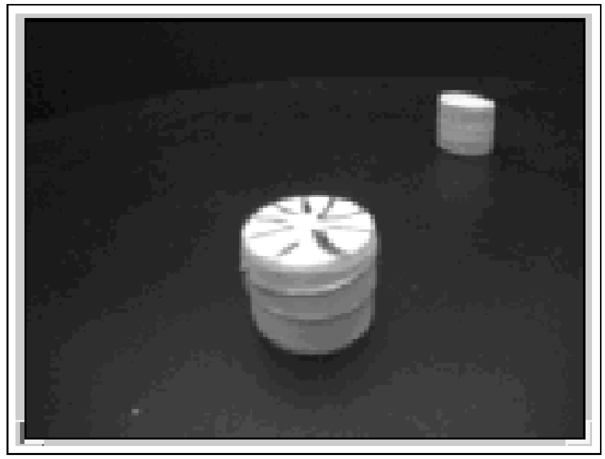

Figure 2: Original image (160x120 pixels) when the robot is in the front of a textured object with a non-textured object in the background.

2.2 Haptic system

The input to the haptic system is as follows. Two wheel encoders provide position information with a resolution of 0.08mm. Arm and gripper position are sensed by position sensors coupled with the respective motors. The arm position sensor takes values from 0 (bottom back) and 255 (bottom forward), the gripper positions sensor takes values from 0 (open) to 255 (closed). Conductivity of objects can be read by a conductivity sensor which takes values from 0 (nonconductive, e.g. plastic, wood) to 255 (conductive, e.g. metal). Objects inside the gripper can be detected by an optical barrier that is mounted on the gripper. The optical barrier takes values from 0 (no object) to 255 (object presence). Finally, there are eight IR sensors, six in the front and two in the back with an angular resolution of 60 degrees each. The maximal distance which the IR sensors can detect is around 40mm. Because of the small distance that they can sense and their arrangement around to body surface of the robot we use them as skin (pressure) sensors.

3 Architecture

3.1 Overview

We first give an overall review of the complete system consisting of a visual and a haptic system linked via reentrant modifiable weights. An overview of the architecture is shown in figure 3. The present model consists of a total of about 2000 units and 140'000 connections. A more detailed description of each individual system is presented below. The main ideas are as follows. We exploit the fact that the agent can interact with its environment in two ways: Firstly, there are focus of attention mechanisms which are directly coupled to the arm, gripper and wheel motors (see below). As a result the agent moves its body

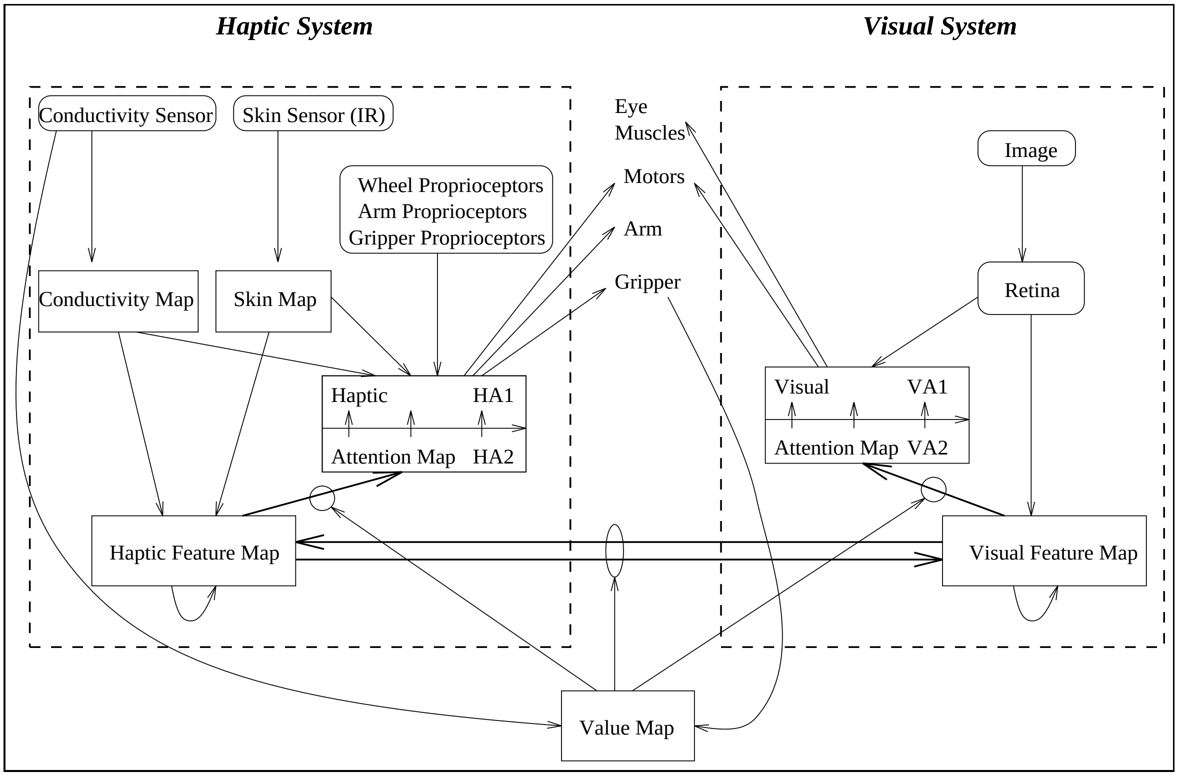

Figure 3: Overview of the architecture. There is a visual and a haptic system. Each system consists of sensory sheets, an attentional map and a feature map. There is a learned crossmodal interaction via reentrant connections between the haptic and visual feature map.

and the arm in a position where it can explore the object. Secondly, instead of categorizing visual information only, both haptic data which result from the exploration process as well as the focused visual data are used for learning and categorizing objects. As shown in figure 3, in both systems there are sensory maps that are connected to feature maps. Feature maps respond to properties of objects such as texture or conductivity. The interaction between the two modalities is implemented via modifiable reentrant weights between these feature maps. The correlation of signals of the haptic and the visual feature map by these reentrant connections forms the basic mechanism of categorization. This is in accordance with the model of categorization in infants proposed by [12]. In sum, the reentrantly connected feature maps form a classification couple ([3]).
A fundamental property of our model is that the feature maps are connected via modifiable feedback connections to the attention maps (see figure 4). The main idea is to link the categorical responses of the classification couple to the attentional sensory-motor loop. In essence, the result of learning is that relevant objects enhance activity in the attentional loop while it is broken in the case of uninteresting objects (which in turn leads to ignoration of the object). Thus, there is no explicit avoidance or approach behavior linked to the classification couple. Instead, the dynamics of the classification couple lead to a differential modulation of the attentional sensory-motor loop depending on the encountered object.

3.2 The visual system

The visual system consists of a retina, a visual attention map and a visual feature map. We use an image of 160x120 pixels as an input to the *retina*. An artifical fovea (20x20 pixels) is extracted from the center of the retina (see figure 5). The spatially averaged periphery of the retina gives input to area VA1 of the *visual attention map*. We have implemented several "attention operators" such as attention to texture, motion or bright spots. In the experiments presented in this paper only the latter operator was used. The visual attention map is part of an attentional sensory-motor loop in two ways. First, it is connected to the wheel motor map. These connections implement a continuous mapping between location of activity in the attention map and the resulting translational and rotational movements. For example, a bright object on the top left of the image will lead to high activation on the corresponding area in the attentional map and this in turn will be translated into forward translation and left rotation. Put shortly, the robot always orients its body towards areas of interest

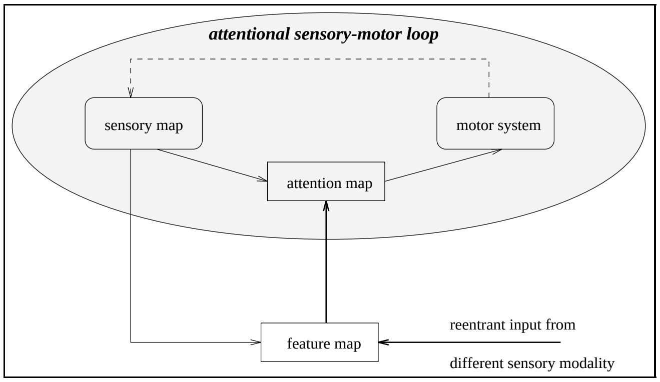

Figure 4: The feedback loop linking categorical responses to the attention system. There is a closed attentional loop between the sensory map, the attention map and the motor system. The resulting sensory-motor coordination is modulated via the modifiable connections between the feature map and the attention map. The main result is that categorical responses are linked in a continuous way to the attentional loop: Relevant objects will enhance activity in the attention map and thus strengthen the sensory-motor coordination. On the other hand, the encounter of uninteresting objects will result in low activations in the attention map and consequently break the attentional loop (sensory-motor coordination).

(bright spots in the present model). In addition, the attention map is connected to artificial eye muscles that move the fovea to bright spots. This saccading leads to a focussing of the fovea on interesting regions in the visual field. Thus, together with the motor map and the eye muscles the attention map forms a complete (visual) attentional sensory-motor loop: it brings the robot to relevant places (as defined by the attention operator) in the environment while at the same time keeping the eye focussed on the spot where the robot is heading to.
The fovea in turn is connected to the *visual feature map*. The feature map responds to relevant features in the fovea. In the present model these are defined to be horizontal and vertical edges, i.e. texture. The visual feature map contains excitatory and inhibitory units; excitatory units make connections to neighboring excitatory units, thus enhancing activity locally. Inhibitory units influence excitatory units further away, thus providing lateral inhibition. This connectivity scheme leads to local and constrained centers of activations. The visual feature map is connected back to area VA2 of the visual attention map via modifiable weights. The main idea behind these weights is that the categorical responses of the classification couple (i.e. the two feature maps) should modulate the attentional sensory-motor loops by either enhancing or breaking it. Area VA2 consists of a population of inhibitory and excitatory units that are connected via prewired weights to area VA1. The activity of these units increases as the weights from the feature map to the attention map evolve. These weights are topographic in the sense that areas that code for strongly textured input are mapped onto the excitatory population, while areas that code for bright or non-textured objects are mapped onto the inhibitory population. High activity of the inhibitory population will lead to a suppression of activity in VA1 and in turn to a breaking of the visual attentional loop. The value map implements a general bias or motivation in the system. The value map receives input from the resistivity sensor and the gripper proprioceptors. The basic motivation behind these connections is that the robot should learn only when it explores an object. Activity in this map acts like a gating function and is used as a reinforcement signal for the synaptic modifications between the feature and attentional map as well as for the reentrant connections between the two feature maps. Finally, the visual feature map is connected to the haptic feature map via reentrant connections. Again, the synaptic growth of these connections is modulated by activity of the value map. The value map is strongly active when the robot has something in the gripper. In this way the robot only learns about objects when it is exploring them with its arm-gripper system. The dynamics of the classification couple, the synaptic growth between the classification couple as well as between the feature maps and the attentional maps will be illustrated in the results section below.

Field	Haptic	Visual
Feature Map	100	100
Attention Map	200	600
Sensory Maps	80(Skin), 100(Conductivity)	700(Retina), 400 (Fovea), 300 (Periphery)
Proprioceptors /Motor Maps	10(Wheels), 10(Arm), 10(Gripper)	
Value Map	20	

Table 1: The number of units per region on both haptic and visual system.

3.3 The haptic system

The overall functionality of the haptic system is similar to the one just described for the visual system. Again, there are sensory maps, an attention map and a feature map. The sensory maps consist of a *conductivity map*, and a *skin map*. They get input from a conductivity and a skin sensor, respectively. Both maps are connected to area HA1 of the *haptic attention map*. Similar to the visual equivalent, the haptic attention map and the wheel, arm and gripper motor maps form an attentional sensory-motor loop. Again there is a continuous mapping

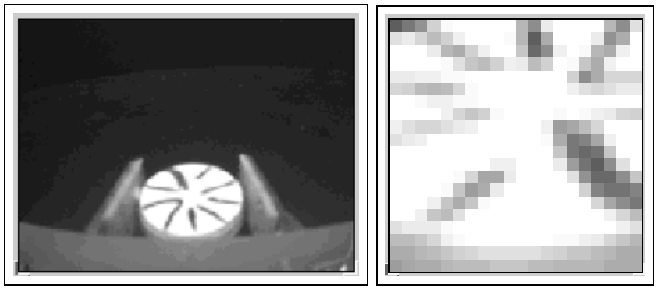

Figure 5: Original image when the robot explores a textured object and the corresponding foveal activity. In this case the foveated region is approximately in the center of the object.

of location of activity in the attentional map and the resulting translational and rotational movements. The main result is that as soon as the robot has made body contact with an object (which is sensed by the skin sensors) it will bring the object to the front of its body. This "haptic focussing" leads to an increase of activity in area HA1 of the attention map and in turn to large activity in the arm motor map, causing the robot to lower it's arm. Lowering the arm leads to increased activity in the arm proprioceptors and in turn to an even more increased activity in HA1. As a result of this the robot will start exploring the object by closing the gripper. Area HA2 consists of inhibitory and excitatory populations of units. They receive inputs from the haptic feature map map and they are fully connected with the units of area HA1. The projections between the haptic feature map and HA2 are topographic in the sense that areas that code for non-conductivity are maped onto the inhibitory population, while areas that code for conductivity are mapped onto the excitatory population. HA2 and HA1 are fully inteconnected. High activity of the inhibitory population will lead to a suppression of activity in HA1. As a result the gripper will be opened and the arm will be lifted. At the beginning of a trial the main source of activity in the haptic attention map stems from the skin map. This is a kind of "haptic reflex" that makes the robot haptically track and explore objects. Over time connections between the haptic feature map and the haptic attention map (area HA2) evolve. This leads to an amplification of attention for relevant objects and to a breaking of the haptic attentional sensory-motor loop when the robot encounters irrelevant objects.

3.4 Activation and learning rules of neuronal fields

The generic equation that describes the activation rule for the attention maps is:

$$a_i^{AM}(t) = \sum_{i=1}^{N^{FM}} a_j^{FM}(t) w_{ij}^{FM}(t) + \sum_{i=1}^{N^s} a_j^s(t) w_{ij}^s(t) \quad (1)$$

where $a_i^{AM}(t)$ is the i-th unit of the attention map, $a_j^{FM}(t)$ is the activation of unit j of the feature map, $w_{ij}^{FM}(t)$ is the weight from unit j in the feature map, $a_j^s(t)$ is the activation of unit j of the sensory map and $w_{ij}^s(t)$ is the weight connecting unit$j$ to unit i. The activity of the units in the feature maps, $a_i^{FM}(t)$, is computed using a "leaky integrator" activation function [1]:

$$a_i^{FM}(t) = \beta a_i^{FM}(t-1) + \alpha\left[\sum_{i=1}^{N^r} a_j^r(t) w_{ij}^r(t) + \sum_{i=1}^{N^s} a_j^s(t) w_{ij}^s(t)\right] \quad (2)$$

where $0 < \beta < 1$ relates to the time constant of the unit, $0 < \alpha < 1$ is the attack parameter, $a_j^r(t)$ is the activation of unit j of the other reentrantly connected feature map, $w_{ij}^r(t)$ denotes the reentrant connection strength from unit j to unit i, $a_j^s(t)$ is the activation of unit j of the sensory map,$w_{ij}^s(t)$ is the weight connecting unit j to unit i. The connections between the sensory map and the feature map are chosen such that they result in a population coding of the average sensory activity (conductivity/skin map activations for the haptic and texture for the visual feature map, respectively). Equation 2 makes the activation of the units not only depend on the weights and the current inputs but also on the activation of the previous time step. Besides being a biologically plausible activation function, the leaky integrator function has the important property that it leads to stability of responses and a kind of "low pass" filtering on the input. The connection strenghts between the haptic and the visual feature map, and between the feature maps and the attention maps are computed using a Hebbian learning scheme:

$$\Delta w_{ij} = v(t)(\eta a_i(t) a_j(t) - \epsilon a_i(t) w_{ij}(t)) \quad (3)$$

where w_{ij} represents the strength of the connection between the presynaptic unit j and the postsynaptic unit i, $v(t)$ is the value signal (activity of the value map), $a_j(t)$ is the activation of the pre-synaptic unit, and η, ϵ are the learning rate and decay parameters respectively.

4 Results

Experiments were conducted on a flat arena (100 cm x 100 cm) with walls (8 cm height) on each side. Objects were of 1.5 cm in diameter and 2 cm high, the shape of the objects was cylindrical (see figure 1). There were

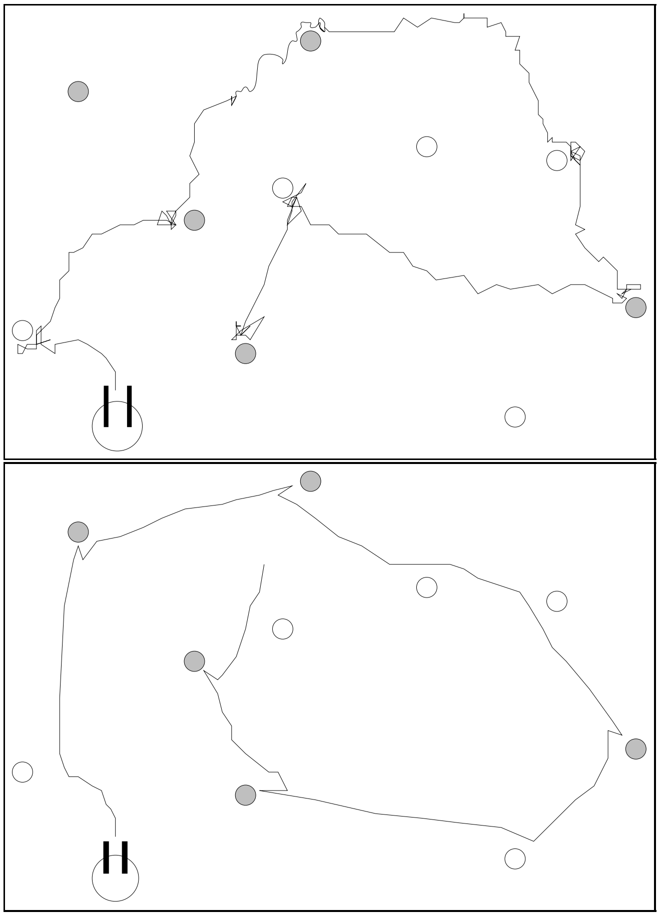

Figure 6: A typical trajectory of the robot before and after learning had occured. For visualisation purposes objects were removed from the gripper once the robot had grasped them.

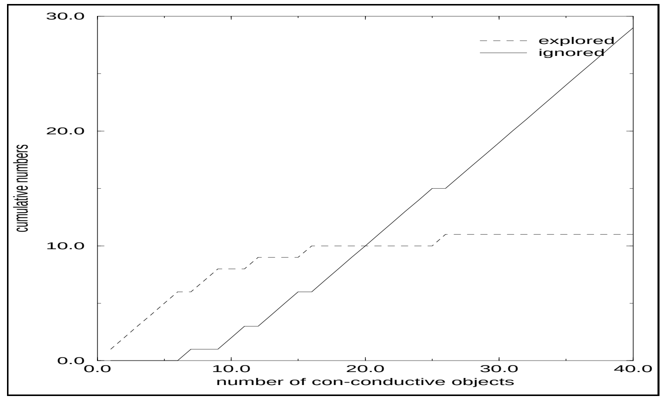

Figure 7: Cumulative number of exploration and ignoration steps for 40 non-conductive objects. The data are means over 20 trials.

conductive and non-conductive objects in the environment. The conductive objects had a strongly textured surface while the non-conductive had a white or only slightly textured surface. The robot's task was to collect the conductive objects. We present results on several levels. First, the overall performance of the system is quantified on the behavioral level. Second, the categorization dynamics are illustrated by measuring the synchronization of activity in the classification couple. Third, the performance and the categorization dynamics are traced back to the dynamics of synaptic change. The behavior of the robot as it moves around in the environment and explores objects is shown in figure 6. The trajectories were recorded with a video camera and then hand traced. Figure 6 (*top*) shows a typical trajectory at the beginning of a trial. White and shaded circles indicate non-resistive/non-textured and resistive/textured objects, respectively. It can be seen there is no distinct behavior for the two types of objects. Rather the robot approaches all objects and explores them. Figure 6 (*bottom*) shows a typical trajectory after the robot has encountered 10 objects of each type. Two main results can be taken from the traces in figure 6. First, the robot has stopped exploring both types of objects. Rather the behavior is now governed by the dynamics of the classification couple. Second, the robot "ignores" non-conductive objects while it grasps the conductive ones (without first exploring them). We use the term "ignoration" instead of "avoidance" to indicate that there is no seperate avoidance module. Rather the avoiding is achieved by breaking the attentional sensory-motor loop. In order to quantify this learning process the number of non-conductive objects explored and ignored was recorded for 20 trials. Each trial ended when the robot had encountered 40 non-conductive objects. Figure 7 shows the averaged cumulative number of non-conductive objects the robot explores and ignores.

At the beginning of the trials the agent always explored non-conductive objects. After it had explored around 6 objects it started to ignore them. Because the weights had not been sufficiently evolved it still explored some of the objects. After having encountered around 12 objects the robot only explored because of errors in the sensory readings. Since the main interest of this paper is on categorization, the dynamics of the classification couple during a typical trial was recorded. One of the main consequences of using reentry as the mechanism for categorization is that temporal correlations or synchronizations in the activity between neuronal areas emerge. In order to quantify these correlations we estimated the level of synchronization between the visual and the haptic feature map. We have quantified synchronization as the product of the average activities of the areas V1, V9 and H1, H9, in the visual and haptic feature maps, respectively. Units in area V1 selectively respond to non-textured regions in the fovea while units in V9 respond to textured ones. Similarly, H1 and H9 of the haptic feature map selectively respond to non-conductive and conduc-

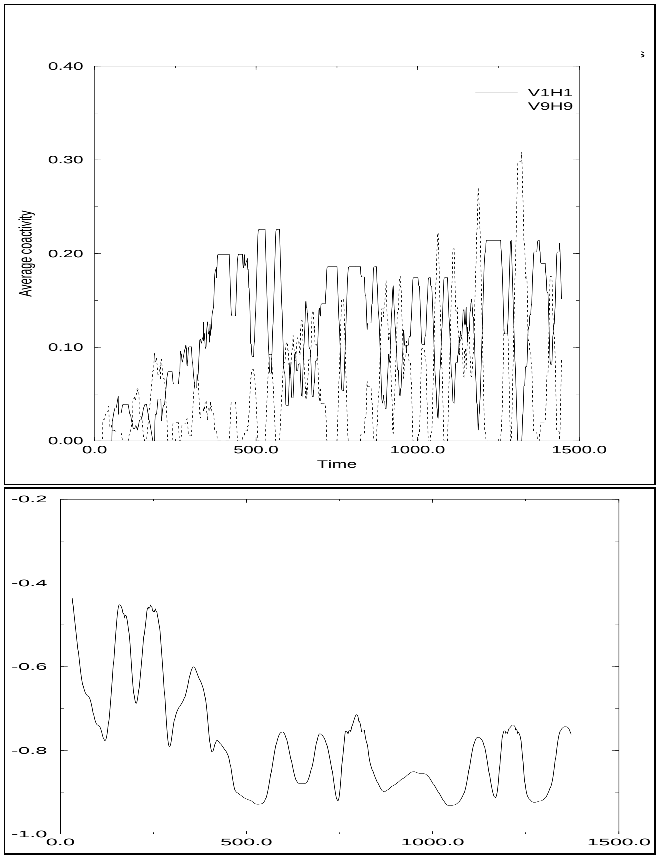

Figure 8: *Top.* Level of synchronization between areas H1, V1 and areas H9, V9. *Bottom.* Correlation of level of synchronization between H1V1 and H9V9. See text for explanation.

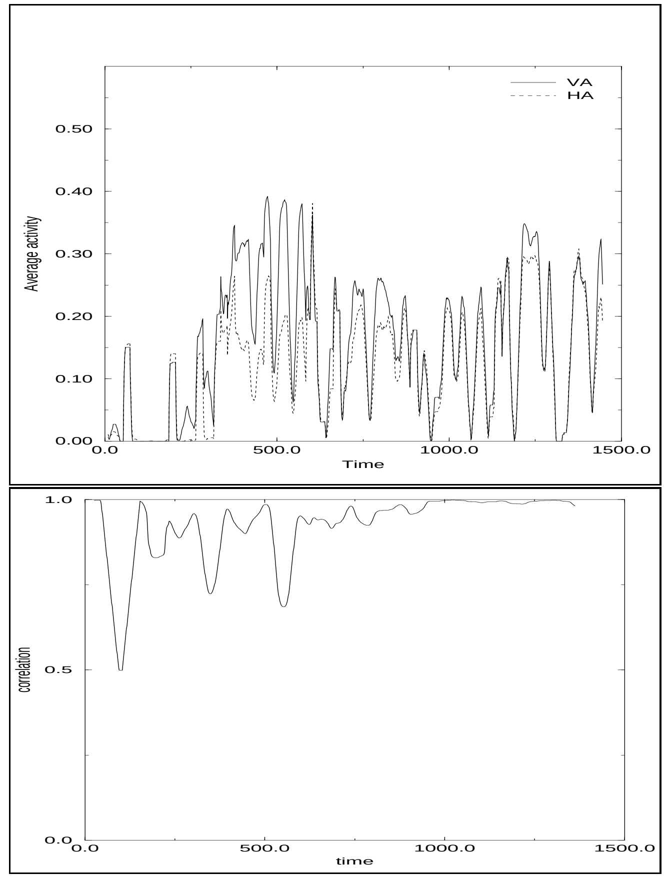

Figure 9: *Top.* Level of synchronization between areas HA2 and VA2 of the haptic and the visual attention map, respectively. *Bottom* Correlation between level of synchronization of HA2 and VA2. See text for explanation.

tive stimuli, respectively. Figure 8 shows the evolution of synchronization between the classification couple as the robot explores objects. Three main results can be taken from this figure. First, the *level of synchronization* between areas V1, V9 and areas H1, H9 (denoted as V1H1 and V9H9 in figure 8, *top*) at the beginning of the trial is significantly lower than after learning has occured. For example, until around 300 steps the mean synchronization level is around 0.05 while later in the trial it increases up to 0.3 implying levels of average activity around 0.55. This increase is due to the evolving reentrant weights that couple the activities in the two feature maps (see below). Second, the *difference* in the synchronized activities between the two areas (V1H1 and V9H9) increases significantly. This can also be seen from the negative correlations that evolve between the coactivities of V1H1 and V9H9 (see figure 8, *bottom*). Third, the *coherence* of synchroneous activity between the maps increases. As can be seen in the figure, the dynamics of the coactivities become strongly coupled over time. Again, this is reflected in the correlations V1H1-V9H9 that reach an average level of around -0.9 (see figure 8, *bottom*).

Another important aspect of our model is that the categorical responses (i.e. the synchronised activities) in the classification couple modulate via modifiable weights the dynamics of the attentional fields. Figure 9 illustrates these couplings for the same trial as the data in figure 8. It depicts the synchronization of the average activity of the inhibitory populations in areas HA2 and VA2 of the haptic and the visual attention map, respectively. Note that although these two areas are not directly connected (see figure 3) a strong synchronization emerges between between the two areas starting at around 300 steps and reaching a maximum correlation of 1.00 at around 1000 steps (see figure 3 *bottom*). This is due to the evolving weights between the feature maps and the attentional maps which leads to a projection of the coupled feature map activities onto the attentional maps. Similar dynamics emerge in the excitatory populations of the attention maps (data not shown).

The activation patterns just described can be traced back to the dynamics of synaptic change between the respec-

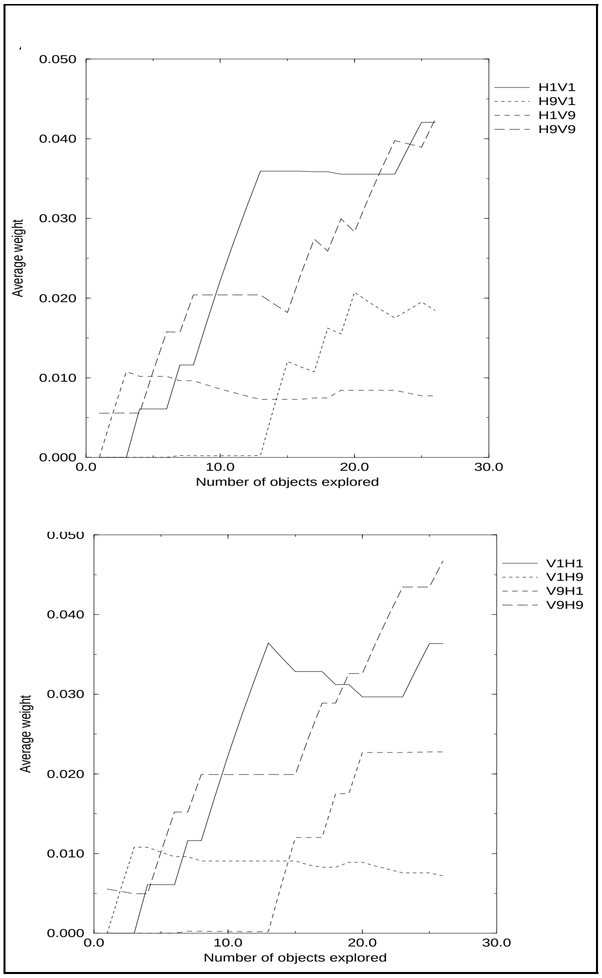

Figure 10: *Top* Average weights from the haptic to the visual feature map and from the visual to the haptic feature map (*Bottom*) for areas V1, V9, H1 and H9 as a function of the number of objects explored. See text for explanation.

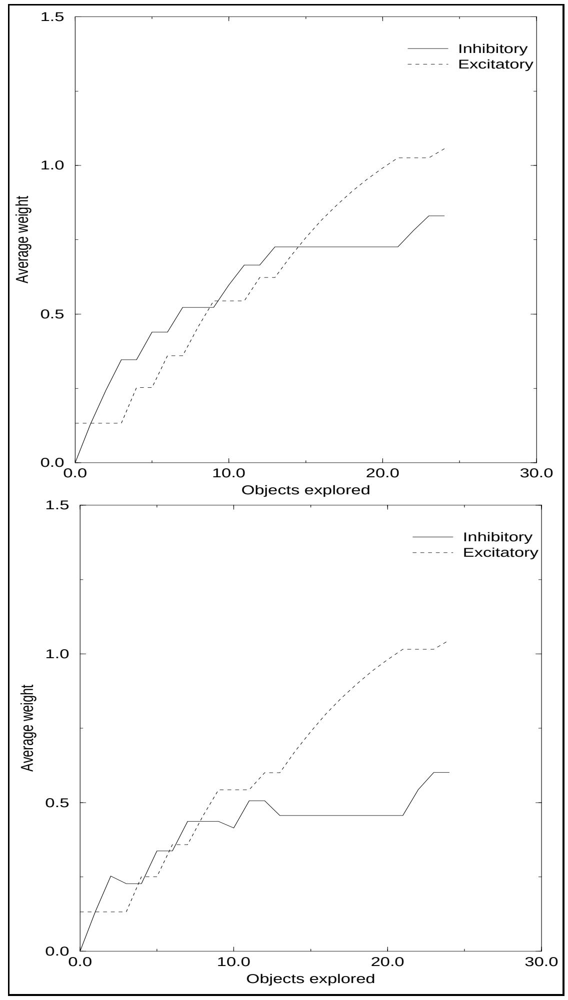

Figure 11: *Top* Average weights from the haptic feature map areas to the haptic attention map and from the visual feature map to the visual attention map (*Bottom*) for excitatory and inhibitory populations. See text for explanation.

tive areas. Figure 10 shows the development of synaptic strength between areas V1, V9 of the visual feature map and areas H1, H9 of the haptic feature map. Again, V1 and V9 denote units tuned to non-textured and textured regions in the fovea, respectively, while H1 and H9 are regions in the haptic feature map tuned to non-conductive and conductive stimuli, respectively. The synaptic strengths between the two feature maps show a similar pattern. Weights between areas responding to stimuli which are reinforced by the value map, i.e. weights between areas V1H1, V9H9, H1V1 and H9V9, increase faster and reach a higher level than the ones which are of no value to the agent. The latter connections (V1H9, V9H1, H1V9 and H9V1) stem from activity in the feature maps which are mainly caused by errors in the sensory maps (i.e. errors in the sensor data). Note that these erroneous weights decrease over time due to the decay term in equation 3. The steepness of the curves in figure 10 is a function of the sequence of objects the agent encounters. For example, it can bee seen in figure 10 that after having explored around eight objects the agent encountered three non-textured and non-resistive objects. As a result, the weights H1V1 and V1H1 show a significant increase during that period.

Finally, figure 10 shows the evolution of the average weights from the feature maps to the attentional maps. These weights couple the categorical responses to the attentional sensory-motor loops. As can bee seen in the figure, weights to both excitatory and inhibitory populations in the attention maps increase significantly with the number of objects the agent has explored. More specifically, area V1 of the visual feature map (responding to non-textured stimuli) becomes associated with the inhibitory population of the attention map (area VA2, see figure 3). Strong weights also evolve between area

V9 (responding to textured stimuli) and the excitatory population of the visual attention map. Together these connections lead to a suppression of activity in the visual attention map for non-textured objects while the encounter of textured objects lead to an enhanced activity. Similar weights evolve between the haptic feature map and the haptic attention map. Area H1 (coding non-resistive input) becomes associated with the inhibitory population of the haptic attention map while area H9 which selectively responds to conductive objects becomes associated with the excitatory population of the attention map. Again, these connections have the consequence that the activity in the haptic attention map gets supressed for non-resistive objects and enhanced for resisitve ones. The weights in figure 10 together with the ones between the feature maps (see figure 10) are responsible for the categorical responses of the agent (see figures 7 and 6).

5 Discussion

In this paper we have addressed the problem of categorization in autonomous agents. We have presented an architecture that is based on the concept of sensory-motor coordination rather than the one of information-processing. Sensory-motor coordination was used on different levels. First, a saccading mechanism was used to focus the fovea on interestings regions in the environment. The sensory-motor coordination here is a reflex based on the fovea and the eye muscles. It leads to a significant decrease of the number of degrees of freedom in the visual space because only the foveated part of the image has to been considered. Second, so-called attentional sensory-motor loops were used. The visual attentional loop caused the robot to orient and move towards objects while the haptic attentional loop resulted in a more fine-tuned focussing of the object with respect to the body of the robot. As a result the robot could explore the object by lowering the arm over the object and closing its gripper. Finally, sensory-motor coordination was involved in the category learning itself. This is because the agent only learned when it had haptically explored an object which in turn is the result of the visual and attentional sensory-motor loops. An important aspect of this is that the categorization is embedded in the interaction of the agent with its environment.
We have shown that based on this architecture the agent is able to learn to categorize the objects in its environment. The categories that evolve are expressed in the synchronization of activity in the classification couple.

References

[1] G. J. Chappell and J. G. Taylor. The temporal kohonen map. *Neural Networks*, 6:441–445, 1993.

[2] M. Dill, R. Wolf, and M. Heisenberg. Visual pattern recognition in drosophila involves retinotopic matching. *Nature*, (365):751–753, 1993.

[3] G. M. Edelman. *Neural Darwinism.* Basic Books, New York, 1987.

[4] R. J. Herrnstein. Levels of categorization. In W. E. Gall G. M. Edelman and W. M. Cowan, editors, *Signal and Sense. Local and Global Order in Perceptual Maps*, pages 365–413. Wiley, New York, 1993.

[5] J. E. Hummel. Object recognition. In M. A. Arbib, editor, *The hand book of brain theory and neural networks*, pages 658–660. MIT Press, Cambridge, MA., 1995.

[6] J. K. Kruschke and M. A. Erickson. Five principles for models of category learning. In Z. Dienes, editor, *Connectionism and Human Learning*, pages 365–413. Oxford University Press, Oxford, 1995.

[7] D. Lambrinos, C. Scheier, and R. Pfeifer. Unsupervised classification of sensory-motor states in a real world artifact using a temporal kohonen map. In *Proceedings of the International Conference in Artificial Neural Networks ICANN'95*, pages 467–472, Paris, October 1995.

[8] R. Pfeifer and C. Scheier. Sensory-motor coordination: the metaphor and beyond. *Journal of Robotics and Autonomous Systems*, in press.

[9] P. Rochat. Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology*, 25:871–884, 1989.

[10] C. Scheier and D. Lambrinos. Adaptive classification in autonomous agents. In *Submitted to EMCSR '96, 13th European Meeting on Cybernetics and Systems Research*, 1996.

[11] C. Scheier and R. Pfeifer. Classification as sensory-motor coordination: a case study on autonomous agents. In *Proceedings of the Third European Conference on Artificial Life ECAL95*, pages 656–667, Granada, Spain, June 1995.

[12] E. Thelen and L. B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action.* Bradford Books, Cambridge, Massachusetts, 1994.

[13] M. Tistarelli. Active/space-variant object recognition. *Image and Vision Computing*, 13(3):215–226, 1995.