Object Category Acquisition based on Physical Agent-Object Interaction

By

Shinya Takamuku

A thesis submitted to Osaka university
for the degree of Doctor of Philosophy
Department of Adaptive Machine Systems
August 2008

Thesis Supervisor:	Minoru Asada
Thesis Committee:	Minoru Asada
	Hiroshi Ishiguro
	Koh Hosoda

Produced in LaTeX 2_ε.

Acknowlegements

The life at Asada lab. and Hosoda lab. was the most fruitful time in my life till today. I cannot thank Prof. Minoru Asada enough for all his supports, guidances and encouragements throughout my research life in the lab. One ideal style of education should be to provide both a great environment to study and the freedom to explore by the students themselves. Prof. Asada has surely provided me with both, and a lot more. I am also very greatful for Dr. Koh Hosoda for his careful and fundamental comments, and especially for showing me how to enjoy research. I hope we have chance to continue some crazy collaboration in the future. I also thank Prof. Hiroshi Ishiguro for teaching me how a fundamental research should be through his tough comments on the thesis. The long discussion at the diffence have greatly changed my thoughts on my research attitude. Many thanks to Dr. Yasutake Takahashi for helping me start my research. RoboCup was an important experience for me to learn the joy of working in a team. Thanks to Dr. Masaki Ogino, Dr. Takashi Takuma, and Dr. Yasunori Tada for taking me out for a drink. Thanks to Dr. Yuichiro Yoshikawa for his careful advices, discussions, and humors. Thanks to Ms. Shizu Okada and Ms. Yayoi Imahashi for all the office duties and the talks. Thanks to Mr. Rodrigo Da Silva Guerra for the philosophical discussions on culture, Mr. Hidenobu Sumioka for his happy smile, Mr. Katsushi Miura for the snacks, Mr. Kenichi Narioka for the coffee time, and Ms. Sawa Fuke for the joyful chats. Special thanks to the collaborators in the lab.; Mr. Tomoki Nishi, Mr. Kentaro Noma, Mr. Atsushi Fukuda, Mr. Takeshi Anma, Mr. Keita Miura, Mr. Shunsuke Sekimoto, Mr. Tomoki Iwase, and Mr. Naoya Yamano. I was happy to work with them. Although I cannot list them all, I wish to thank all the people I met at Asada lab. Some other thanks to the people I met outside the lab also. Thanks to the members of Sony

Intelligence Dynamics lab. for the wonderful time there. I've heard that they still continue the research of Intelligence Dynamics, and hope that the work will prove fruitful in the near future. Thanks to Prof. Ronald Arkin for his careful guidance of the research there and also the exciting discussions at London. Thanks to Dr. Gabriel Gomez for the careful support on my research at AI lab., Zurich. And finally, I would like to thank my parents and brothers for all the support throughout my student life.

July, 2008

Shinya Takamuku

Abstract

The issue of acquiring daily object categories shared within the human society is a fundamental problem of cognition expected to be solved for the development of humanoid robots. Computer vision studies have started to make progress in solving the problem, but unsupervised learning of the diverse daily object categories is a difficult problem without any conclusive solution. Here, we give an alternative approach to the issue suggesting that category acquisition based only on images and movies is simply an ill-posed problem, and that physical agent-object interaction plays an essential role in solving the problem. The disseration consists of two parts; primitive category acquisition based on infant-like touch and lexicon acquisition based on behavior learning. The former part reproduces infants' typical physical interaction with an anthropomorphic robot to show how the body wisely extracts information which specify the categories through those interactions. This part consists of a work which shows that the very first manual touches observed in infants are capable of distinguishing surface stiffness of objects based on static and dynamic deformation of the skin, and another work showing that shaking behaviors enable acquisition of primitive categories such as rigid objects, liquid, and paper materials, based on auditory information processing of the cochlea. The latter part on the other hand approaches the task of lexicon acquisition by learning object-oriented behaviors shared within the category. By introducing a multi-module reinforcement learning, the robot can acquire and identify the behaviors which specify the category to generalize words to new objects with similar affordances. While the former part deals with low level categories such as those related to materials, the latter part deals with higher level categories related to functions such as those of tools.

Contents

1 Introduction **1**

1.1 The issue of object categorization 1

1.2 Different field of studies on the issue 2

1.3 Physical agent-object interaction for categorization 3

1.4 Overview 4

2 Related works on object categorization **7**

2.1 Computer vision based approach 7

2.1.1 Visual feature based categorization 8

2.1.2 Social cue based categorization 11

2.2 Physical interaction based approach 12

2.2.1 Multimodal information based categorization 12

2.2.2 Sensorimotor coordination based categorization 14

2.3 Summary 18

3 Primitive object categorization based on infant-like touch **21**

3.1 Introduction 21

3.2 Early manual touch experiment 21

3.2.1 Robot setup 22

3.2.2 Robot task 25

3.2.3 Exploratory procedures 25

3.2.4 Results 28

3.2.5 Conclusion 32

3.3 Dynamic touch experiment . . . 33
3.3.1 Related work . . . 34
3.3.2 System . . . 34
3.3.3 Experiment Design . . . 35
3.3.4 Results . . . 37
3.3.5 Conclusion . . . 41

4 Lexicon acquisition based on behavior learning 43
4.1 Introduction . . . 43
4.2 Lexicon Acquisition based on Behavior Learning . . . 44
4.2.1 Basic Idea . . . 44
4.2.2 System Overview . . . 45
4.2.3 Learning and identifying object-oriented behaviors . . . 47
4.2.4 Categorization of photometric feature space . . . 50
4.2.5 Learning relation between object-oriented behaviors and labels . . . 52
4.2.6 Label association policy . . . 52
4.3 Experiments . . . 54
4.3.1 Task . . . 54
4.3.2 Learner Setup . . . 56
4.3.3 Experiment on object-oriented behavior learning and identification . . . 58
4.3.4 Experiments on learning relation between object-oriented behaviors and labels . . . 60
4.3.5 Experiment on label association . . . 61
4.4 Conclusion . . . 62

5 Conclusion and future work 63

References 67

Chapter 1

Introduction

1.1 The issue of object categorization

A challenge to develop a humanoid capable of helping human users in daily environments is now becoming real [1] [2] [3] [4]. Watching entertainment robots in houses [5], cleaning robots in officies [6], and surgical robots in hospitals [7], it seems that the dreamed future of robots coexisting with humans is about to come. However, as we have more and more opportunities to meet such robots, we soon notice that what the robots can do is still limited. Most robots developed till today can detect only limited kinds of objects such as human faces, colored balls, and markers. They will fail to detect the objects by slight changes in texture, size, shape, material, lighting conditions, backgrounds, and so on. Hence, a household robot will not be able to hand the user an unfamiliar magazine when asked for it, cleaning robots will fail to separte burnable from non-burnable trash, and rescue robots will continue to approach unstable places at disaster sites. Current robots lack the most basic cognitive ability to deal with the increasingly complex environment; the ability to identify categories. A *category* is a group of objects, actions, or environmental states which have similar meanings to an agent or a group of agents. For example, various magazines with different design, size, and paper materials are all considered to be an object from an identical category by humans. Organization of experiences through the use of such categories dramatically reduce the complexity of the outer world, enabling them to learn appropriate

behaviors and to communicate with other agents by exchanging labels corresponding to some shared categories. Since categorization underlie most of the higher-level cognitive abilities, the understanding of categorization processes is essential for fundamental study of cognition.

This research address the issue of acquiring daily object categories shared within the human society. Current robots are often equipped with hand coded recognition modules for fixed set of object categories. Considering the diverse and increasing objects in daily environments, robots should ideally obtain the ability to acquire object categories by themselves through their experiences. The topic is crucial for robots designed to coexist with human users in daily environments for two main reasons. Firstly, since the human categories are tuned for living in the daily environment, it is likely that the categories are also useful for the robots to work in the same environment. Secondly, sharing categories is a basis for natural communication [8]. Obtaining humanlike categories will enable the robot to view the world in a manner similar to that of humans, and form the basis for lexicon acquisition leading the way to language communication.

1.2 Different field of studies on the issue

Inspite of the importance of the issue, this acquisition of humanlike categories is still one of the most difficult problems of cognition. Since the category acquisition should intrinsically handle the diverse set of complex objects we face in our everyday life, it is difficult to formalize the task and simulations or any other purely symbolic approaches [9] [10] [11] involve the possibility of being impractical. Once the task is formalized into some artificial problems with abstract objects such as colored spheres, it is no longer clear whether we are still facing the same problem. Another approach to address the issue is observation of humans who manage with the same problem. In fact, the nature of human categories have been investigated by philosophers, psychologists, and linguistics for centuries [12]. The mystery how human infants learn meanings of words from few experience [13], known as the Gavagaii problem [14], have inspired the linguistic studies of human categories. By observing the generalization of unfamiliar words to new objects, they have proposed innate biases on categorization [15] and the fact that shapes [16], functions [17], and behaviors of adults [18]

can be the keys of the categorization. Recently, neuroscientists have joined the study finding category specific activity in the human medial temporal lobe [19]. Developmental scientists also joined the study by introducing approaches such as familialization methods [20]. These observation studies have revealed the nature of human categories to some extent. However, when we try to investigate the mechanism of acquiring those categories, we face a critical problem. Current measuring techniques including brain imaging methods do not allow us to investigate the whole process of categorization from brain activities to behaviors [21]. Even though we have such techniques, the problem of observing daily processes remains. The use of daily objects in observation experiments involves effects from daily life of the infants which makes it difficult to investigate the pure process of acquiring those categories.

Since there isn't a complete theory for object category acquisition, engineers are asked to build artificial systems capable of acquiring humanlike categories through trial and error. This challenge however, turns out to be a approach which overcomes the problems of the observational or theoretical approaches. By introducing robots, we can introduce daily objects to the experiment and systematically investigate the body setup, control, environmental design, as well as the information processing required for the categorization process.

1.3 Physical agent-object interaction for categorization

One obvious difference of current robots and infants is the fact that robots are often just passively observing the objects with their camera, whereas infants actively explore the objects through physical interaction to obtain multimodal representation of the objects [22]. Several findings from various fields of study indicate the possiblity that this absense of physical interaction with the object is the fundamental cause of the low performance of current robots in object categorization. Psycholinguistic studies show that the first categories obtained by infants are based on physical experiences [12], and various daily object categories also seems to be the case. Categories of letters for example are found to be related to the action of writing them [23]. It is also well known that function is an important information for human categories as indicated in the word generalization experiments [17]. Note that such categories related to functions cannot be determined by pure shape, but should take in

to consideration the action to use them. Recent brain imaging studies indicating the role of action modules in categorization of manipulative objects [24] also supports the idea. The increasing evidences indicating the role of physical interaction in obtaining humanlike categories calls for a theory to explain them. The theory of information pickup [25] by J. J. Gibson provides a general explanation why physical interaction is required for object category acquisition. A big challenge for object category acquisition is to acquire categories shared within the agents so that they could be used for communication. If the categories are built in an arbitrary manner, agents will not reach such a shared cognitive structure. Instead, they should have some common basis for it. Gibson's idea is that the shared environment plays a role in this aspect, affording the agents similar information structures as invariance in sensorimotor experiences when the agents engage in intense physical interaction with the environment. The idea is advanced by several researchers such as E. Gibson insisting the role of exploratory behaviors in categorization [26]. Thelen and Smith [27], on the other hand, proposed the idea that categories can self-organize through multimodal correlations in real time, along with evidence by simulations on letter categorization [28]. Pfeifer and Scheier [29] further showed the importance of sensorimotor coordination in categorization through robot experiments. Their claim is that coordinated action toward the objects is required to transform the ill-posed problem of object categorization to a solvable one reducing the sensori-motor space into a manageable scale.

1.4 Overview

Although there are several general theories of object categorization based on active exploration, the mechanism behind the process of acquiring humanlike daily object categories through physical interaction is not known. Theoretical approaches and observational approaches will face great difficulties when considering this physical categorization due to its complexity and difficulty of objective measurement. Accordingly, this dissertation aims to clarify the mechanism through the approach of cognitive developmental robotics [30], that is, to understand the mechanism by buiding robots capable of such categorization. The dissertation, in particular, address two possible ways of physical interaction taking a role in

object categorization. The contents are the following:

Primitive object categorization based on infant-like touch

In chapter 3, we introduce the work of reproducing infant-like exploratory behaviors with anthropomorphic robots, clarifying how the body extracts information to specify primitive object categories of daily life. The first part of this chapter investigates the initial manual touch of infants with skin-covered robotic hand. The result shows that the behaviors, squeezing and tapping, could distinguish various surface conditions by the static and dynamic deformation of the manual skin. The second part of the chapter, on the other hand, implements the shaking behavior of infants, refered to as dynamic touch. Here experimental results show that the dynamic touch enables acquisition of primitive categories such as rigid objects, bottles with liquid, and paper materials, by auditory information processing of the cochlea.

Lexicon acquisition based on behavior learning

In chapter 4, we address the problem of lexicon acquisition by introducing learning of object-oriented behaviors shared within the category. The idea is to learn object categories of words not by finding direct correspondence between words and visual features of the objects, but instead, by learning the correspondence between words and the behaviors afforded by the objects of the word category. The proposed architecture could be considered as the model of infants' word generalization based on functions [17]. Experimental result are given to show that the robot can generalize words to unfamiliar objects by identifying the behaviors afforded by the object.

While the former part deals with low level categories such as those related to materials, the latter part deals with higher level categories related to functions such as those of tools. Although human handle various kinds of daily object categories including abstract ones such as "vehicle" and "food", the dissertation address the issue of obtaining the basic level categories grounded to the physical experience.

Chapter 2

Related works on object categorization

One long standing issue yet to be answered is whether physical interaction is essential for object categorization or not. To consider this issue, I will introduce two schools of studies on object categorization; computer vision based approach, and physical interaction based approach. The former is the most popular field of study on object categorization examining how object categories can be acquired or recognized from images or movies. The latter on the other hand shows how much physical interaction can play a role in object categorization through bodily sense, active exploration, and sensorimotor coordination. Note that the two approaches do not conflict to each other, but can be combined. The overview of existing works of each approach follows and a summary ends the chapter.

2.1 Computer vision based approach

Object categorization, also refered to as 'generic object recognition', has become a popular topic in the field of computer vision from the recent success in recognition from local salient parts [31]. In this section, we will first describe the basic procedure for visual object categorization, followed by some modified methods utilizing social cues such as humans' speeches and actions.

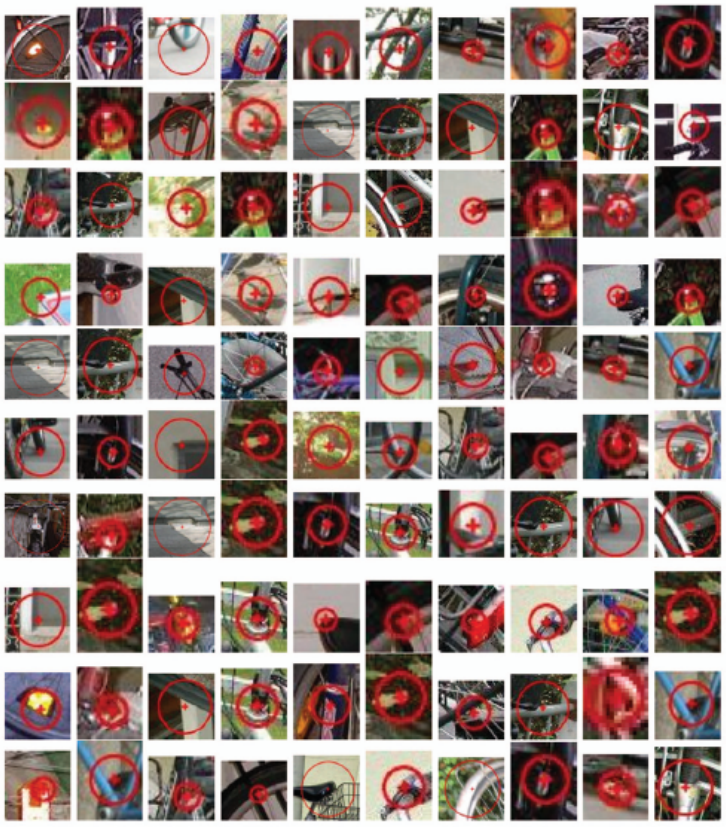
2.1.1 Visual feature based categorization

The task of object categorization is to identify the category of a object in an image or a sequence of images along with their location [1]. A typical procedure for object categorization is as follows. First salient points in the image is detected. Salient points are defined as locations in the image with significant changes in more than one direction and various detectors are invented [32] [33] [34] [35] [36] [37]. Affine invariant detectors [38] [39] are also proposed to manage variety in scale and pose of objects. Then, feature vectors for each image event are calculated from local description of the detected salient points. The most simplest method to model an object category based on these descriptions is the 'bags of keypoints' approach [40] [41]. This approach represents an object category as sets of descriptions of salient points, or clusters of them. An example of the model for a bike in GRAZ02 database is given in Fig.2.1. Another popular method to model an object category is the constellation approach [42] [43]. In this method, spatial relationship between parts are also included in the representation. Faces for example can be recognized by detecting eyes, noses, and mouthes and considering their spacial relationships. Finally, given the feature vectors, object categories can be learned and classified by utilizing probability based methods such as *maximum likelihood* and the *Bayesian* parameter estimation.

Although the methods described so far requires supervision for learning the categories, obtaining a sufficient number of suitable visual data with annotations is not a easy task. There are recent proposals to utilize the massive amount of image data online such as Google image search [44]. However, such search results contain many outliers, and careful training procedures are required [45]. Consequently, unsupervised category learning methods are also proposed, and proved to be efficient at least for the popular categories to be investigated with visual categorization such as bikes, cars, airplanes, and human faces [43] [46].

[1] It can also be the case that objects from learned categories are not included in the image.

(a) example bike image (b) Harris Laplace keypoints



(c) 'bag of 100 keypoints' model

Figure 2.1: A 'bag of keypoints' model for bikes from the GRAZ02 database [41].

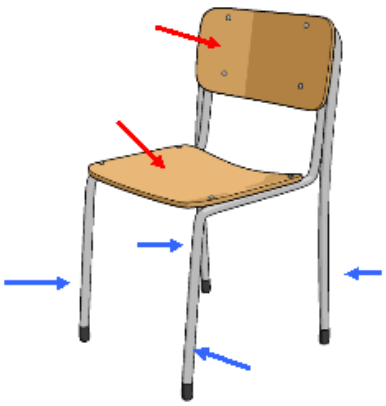


FIGURE 2.2: Features relative to the category chair cannot be specified from vision.

Recent development of visual categorization methods have shown a great increase in their performance. However, the performance is measured with only limited kinds of categories such as bikes, cars, airplanes, and human, provided in the popular benchmark datasets of ETH-80 [2], Caltech [3], UIUC [4], TU Darmstadt [5], and TU Graz (GRAZ01 and GRAZ02) databases [6]. Considering the fact that the appearance of objects is not equal to the ecological meaning of them, the performance which visual categorization can reach seems limited. Imagine classifying steel cans and aluminum cans or empty cans and filled cans. Such difference is important in an ecological sense, but extremely difficult to determine from pure image. We could also imagine obtaining a category of tools like chairs. An agent will not be able to specify which part is relative to the category of a chair from an image alone. In summary, visual information should be useful for recognition of categories, since humans can also recognize objects by sight. However, since similarities in local shapes and textures do not necessarily represent the similarities of objects' meanings to humans, it is questionable whether the approach could solve the problem of "acquiring" the huge set of daily object categories without vast amount of teaching.

[2]http://www.vision.ethz.ch/projects/categorization/eth80 - db.html
[3]http://www.vision.caltech.edu/html - files/archive.html or http://www.robots.ox.ac.uk/ vgg/data3.html
[4]http://l2r.cs.uiuc.edu/ cogcomp/Data/Car/
[5]http://www.pascal - network.org/challenges/VOC/databases.html
[6]http://www.emt.tugraz.at/ pinz/data

2.1.2 Social cue based categorization

One major solution for facing the difficulty of acquiring object categories through vision is to rely on social cues. The idea is to group the objects into categories based on the similarities of behaviors of humans toward those objects. One natural example of such approach is found in the lexicon acquisition system of Roy and Pentland [47]. Their system learns lexicons by statistically modeling consistent cross-modal structure between videos which capture a object and infant-directed speech about the captured object (Fig.2.3).

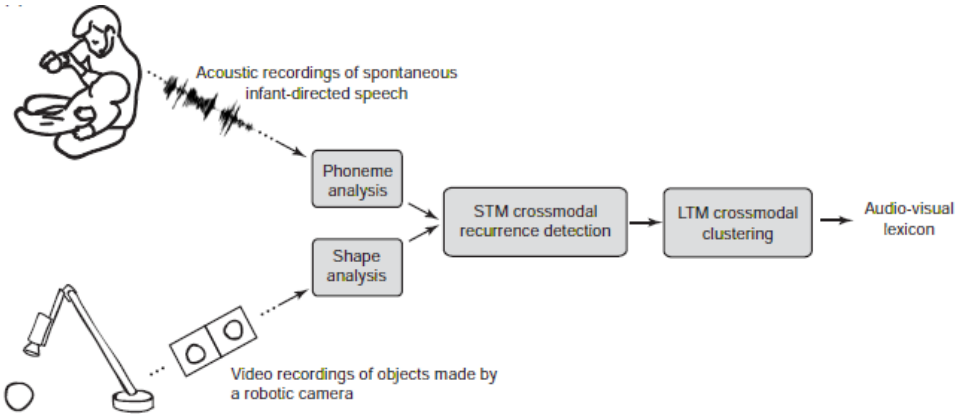


Figure 2.3: The model of lexicon acquisition by Roy and Pentland [47].

Another popular method of this approach is visual object categorization based on motions of humans in videos. Psychological evidence found by Kobayashi et al. [18] that the caregiver's actions toward the object affect infant's generalization of words to new objects is consistent with this approach. Indead, important information which specify the category such as functions can be indirectly obtained by observing humans facing the object. Moore et al. [48] proposed a method of object classification by exploiting human motion. In their method, Bayesian classifiers are utilized to label actions represented as HMMs, which are subsequently utilized to classify the objects. Similar idea is also employed in the FOCUS (Finding Object Classification through Use and Structure) system proposed by Veloso et al. [49]. An example result of the FOCUS system detecting chairs is shown in Fig.2.4. Although these method of exploiting human behaviors could introduce the information of affordance to the categories, there are still difficult issues left to be solved for their implementation. It

could be claimed that the problem of object categorization is just swapped into the problem of action categorization, which is also a critically difficult problem without any conclusive solutions. We should also note that the categories obtained from observation are not always useful for the observing agent due to the differences of motor abilities and body structures.

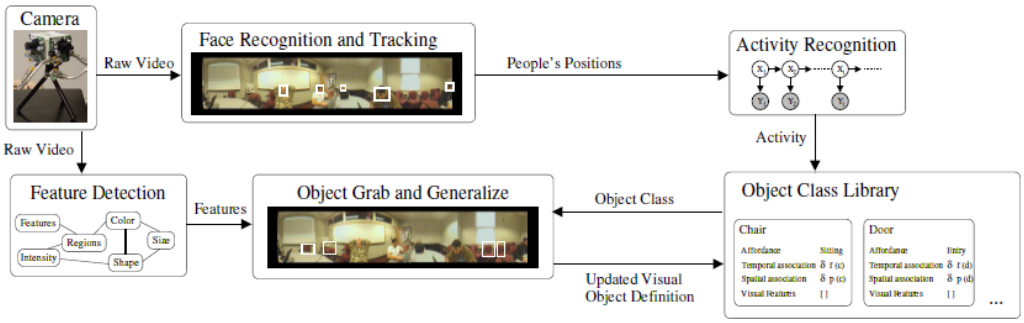


FIGURE 2.4: The FOCUS method [49] detecting chairs in the environment.

2.2 Physical interaction based approach

As discussed in the introduction, the role of physical interaction for object categorization has been insisted in various fields of studies such as psychology, linguistics, brain science, and robotics. In this section, I will give an overview of previous studies related to this topic in the field of engineering. First, the approach of object categorization based on multimodal sensory information will be described, followed by further approaches considering sensorimotor coordination.

2.2.1 Multimodal information based categorization

The finding that human infants, who are experts of object categorization, develop multi-modal representation of objects through physical interaction with them [22] have inspired the development of artificial systems capable of such ability. By physically interacting with the objects through self-produced actions, agents can obtain multi-modal sensory information through tactile, kinesthetic, gustatory and auditory senses. Considering the statement

of J. Gibson [25] and E. Gibson [26], invariant properties of the objects obtained from such physical interaction could be the source of shared categories due to its objectivity. To be more precise, such senses obtained from physical interaction reflect the physical properties of the objects often related to the use of them [7]. Consequently, it is often the case that the key features to bound the objects into categories can only be found in such senses obtained through physical interaction.

Several attempts are made to reproduce the infants' active exploration behaviors toward the objects with robots to investigate the role of them in object categorization. Manual exploration is among all the most intensely investigated exploratory behavior. The exploratory procedures [50] used by adults to explore haptic properties of objects are lateral motion (a rubbing action) for detecting texture; pressure (squeezing or poking) for encoding hardness; static contact for temperature; lifting to perceive the weight; enclosure for volume and gross contour information; and contour following for precise contour information as well as global shape as shown in Fig.2.5. It is pointed out by Bushnell and Boudreau [51] that the ages of acquiring these procedures are consistent with the ages of recognizing the features in human development. Such exploratory behaviors are implemented and investingated in the robotic field. The role of lateral motion on detecting texture for example, is investigated by Tada et al. [52] with their anthropomorphic finger.

The researchers of the RobotCub project[8] have put a lot of effort to reproduce the manual explorative behaviors of infants with their robots. Natale and his colleagues [53] developed a robot with an elastic joint hand (Fig.2.6 (a)) to obtain physical representation of objects through grasping. In their experiment the robot sensed the shape of the grasping hand with constant torque on the joints to detected the shape and stiffness of the objects. The recognized objects and the clustering of the self-organizing map for each objects are shown in Figs.2.6 (b) and (c). Further experiments are made by adding tactile sensors to the hand of a similar robot to investigate the role of cutaneous sense [54]. Tapping behaviors are also investigated by the robot, utilizing the produced sound for recognizing the object [55].

Dynamic properties of the objects are also investigated. The work by Ogata et al. [56]

[7] For example, an infant can determine if the object is eatable or not by having a bite on it.
[8] http://www.robotcub.org/

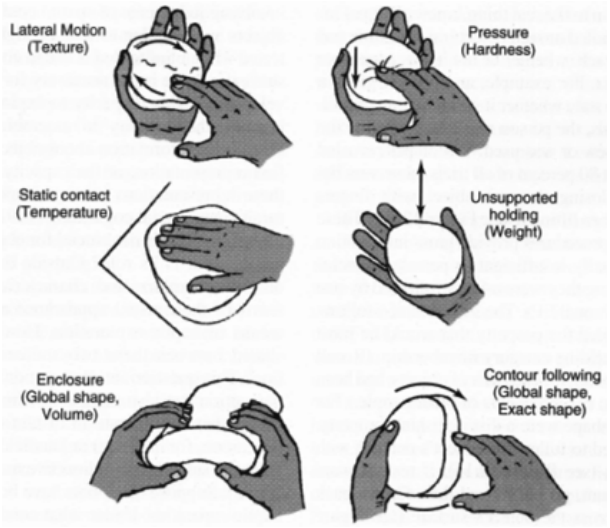


FIGURE 2.5: Exploratory procedures proposed in [50].

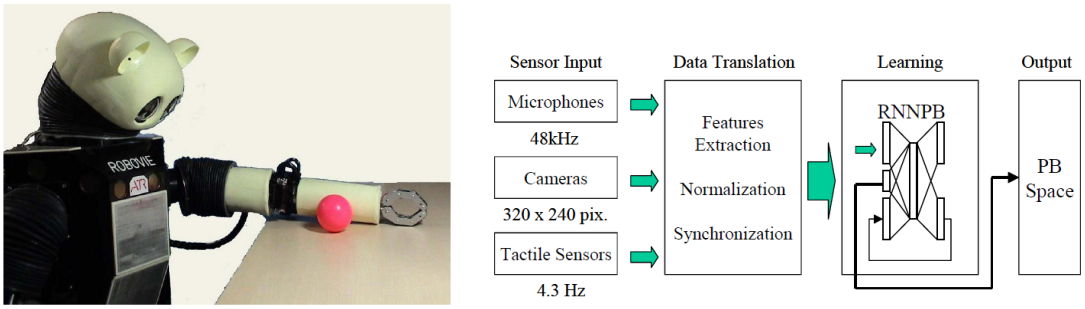
investigating how poking behaviors could form object representations through visual, auditory, and tactile sensory values (Figs.2.7(a)). They apply the RNNPB (Recurrent Neural Network with Parametric Bias) method for representing the dynamic property of the objects (Fig.2.7(b)). Shaking behaviors are implemented by Atkeson et al. [57] and Suzuki et al. [58], showing how the robot could sense the shape of rigid objects through torque; the sensing of momentum of inertia. Finally, object categorization method using audio-visual and haptic information based on probabilistic Latent Semantic Analysis(pLSA) is proposed by Nagai and Iwahashi [59]. There work shows how the addition of auditory and haptic information could ease the problem of unsupervised categorization.

2.2.2 Sensorimotor coordination based categorization

The importance of sensorimotor coordination on object categorization is carefully discussed by Pfeifer and Scheier in their book "Understanding Intelligence" [29], but let us follow and extend the discussion here. Although it was shown by the variety of works on active exploration that self-produced actions could ease the task of object categorization, most works implemented fixed actions carefully modified by the designer. However, the role of actions in object categorization has another aspect of organizing the sensory information to

(a) Babybot: The robot used in the experiment (b) Objects used in the experiment (c) SOM distribution

Figure 2.6: Robot experiment by Natale et al. [53] on grasping behavior.



(a) The robot used in the experiment (b) System of object categorization

Figure 2.7: Poking robot experiment by Ogata et al. [56].

ease the categorization. One could, to start with, refer to the work of Nolfi [60] [61] and Beer [62], having the agent acquire behaviors to specify the category through evolutionaly methods. In the experiment of Beer, the agent producing a number of rays with which it can measure distances from objects was to discriminate between circles and diamonds falling from above as shown in Fig.2.8. The agent could move horizontally and the neural network which controlled the agent was evolved using a genetic algorithm. The optimal policy for the discrimination obtained was as follows: The agent first centered the object and then actively scanned the object until time to make the decision. In this case, the agent is selecting and reducing the sensor space to determine the category by forming a standard position with

respect to the object. The result is analogous to the finding that infants often explore objects by moving them in front of their face to normalize the size [51].

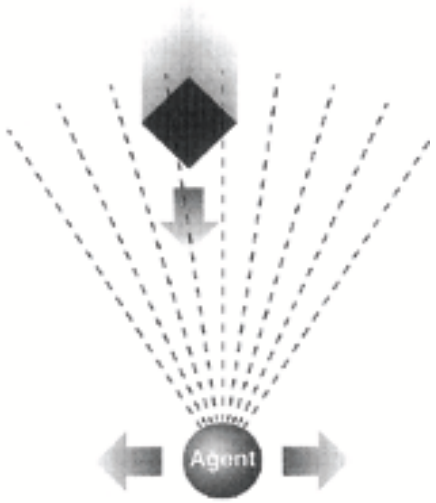


FIGURE 2.8: Setup of Beers experiment [62].

Another example of sensorimotor coordination taking a role in object categorization is given as the experiment of SMC agent by Scheier and Pfeifer [63]. In their experiment, the agent with encoders on the wheels and eight IR sensors had to distinguish cylinders with different size. Due to the poor sensor provided, the task was a non-trivial one. However, the agent managed to distinguish the objects by circuling around the object and utilizing the sensory sequence obtained during this behavior as shown in Figs.2.9. In this case, sensorimotor coordination is again serving as to form the sensory space suitable for the categorization.

Finally, we introduce the series of works by Edelman and his colleagues on categorization. Edelman claims that categories self-organizes from the activity dependant reentrant mappings of multiple disjunctive processes that operate over the same input in real time. A system of letter classification called Darwin II [28], designed by this idea is shown in Fig.2.10. Darwin II consists of two systems called "Darwin" and "Wallace", respectively. Each system are organized of two layers, both receiving input from the same input array.

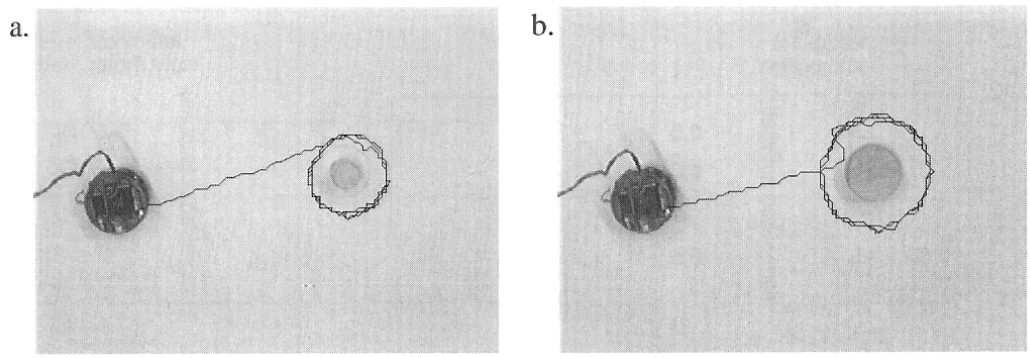


Figure 2.9: SMC(Sensori-Motor Coordination) agent sensing the size of a circule by circling around the object [63].

The function of "Darwin" is to extract features such as lines, orientations, and curves of incoming letters. The R layer consists of a number of identical feature analyzing network, each of which extracts the same features such as lines, orientation, line terminations, etc. Then, the R-of-R layer with numerous random connections to R extracts the nonlocal combinations of the local features of R. The function of "Wallace", on the other hand, is to map the input letter to the movement sequences of a continuous tracing of the letter. The Trace layer extracts the sensory input of a finger tracing object independent of translations and rotations. Then the higher layer responds to the combination of the activity in Trace layer to obtain the global representation of the letter. Finally, the correlations of the activity of the two higher layers are obtained by Hebbian Learning to teach each other the categories of the letters. The system shows how the view of the world as an actor could affect the categorization process. The biological plausibility of the result is supported by observation study of human letter categorization showing that classification of newly learned letters are affected by the taught drawing techniques of them [23].

In summary, action is not only important for obtaining multi-modal representation of objects, but also plays a role easing the task of object categorization by reducing the sensorimotor space into a manageable scale. Furthermore, the idea that actions which are afforded by the objects could be the key information for categorization of humanlike categories was introduced.

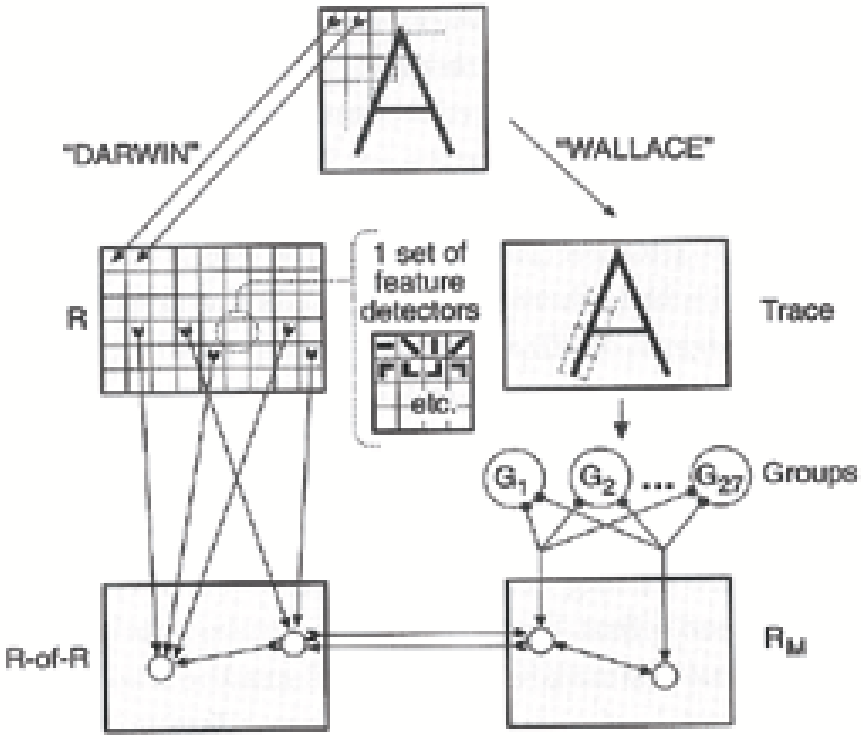


FIGURE 2.10: An overview of Darwin II capable of learning letter categories [28].

2.3 Summary

To summarize the chapter, I would draw an overall sketch of the process of categorization through physical interaction. When an agent is given an object, they will actively explore it with their body to obtain multimodal information of it. By modifying the actions toward the objects, in some cases by focusing on the characteristic aspect of them, he or she will obtain a reduced sensorimotor space suitable for object categorization. And once the categories are acquired through the interaction, visual information of the objects could be utilized to recognize them.

While existing works have shown the general theory of object categorization, two factors which seems essential for acquiring "humanlike" object categories remain not addressed. First key factor is embodiment. According to the evolution theories, the design of the human body is optimized to extract information essential for their survival from the environment. It is then natural to consider the human bodies to play a considerable role in object categorization, and artificial agents may also benefit from humanlike body design in obtaining humanlike categories. The second key factor is affordance. Many researchers have pointed out that functions account for the major part of human categories. This indicates that actions are not only for organizing the sensory space for categorization, but actions afforded by the objects are itself an important property for object categorization of humans. However, general theory on how such affordance could be utilized for object category acquisition is not shown.

The dissertation address the issue on how the two key factors, embodiment and affordance, plays a role in acquiring humanlike categories. Chapter 3 address the issue of obtaining primitive categories through infant-like touch. The work focuses on how the body extracts information important for object categorization through typical physical interaction. Chapter 4 on the other hand address the mechanism of object categorization for lexicon acquisition based on behavior learning. The latter work tries to explain how object affordance could be autonomously obtained and utilized for categorization.

Chapter 3

Primitive object categorization based on infant-like touch

3.1 Introduction

While robots are often only passively observing the objects, human infants acquire multimodal representation of them through physical exploration [22]. As stated in the previous chapters, such interaction toward the objects plays a fundamental role in object categorization. Since the aimed task is to acquire the categories shared within the human society and that acquired categories depend on the behaviors performed by the agent [28] [23] and its body, robots may also be able to acquire humanlike categories by imitating the exploratory behaviors of infants with anthropomorphic body structure. Here, we introduce two robotic experiments based on such idea; early manual touch experiment with skin covered robotic hand and dynamic touch experiment with pneumatic robot arm. Both work shows how the body extracts information relavant for categorization through the physical interaction.

3.2 Early manual touch experiment

The sense of touch plays an important part of object representation in the early stage of development [64]. Lederman and Klatzky [50] suggested the existence of special exploratory

behaviors called exploratory procedures which enables the agent to detect haptic properties of objects; lateral motion (a rubbing action) for detecting texture, pressure (squeezing or poking) for encoding hardness, static contact for temperature, lifting to perceive weight, enclosure for volume and gross contour information and contour following for precise global shape. It was then pointed out by Bushnell and Boudreau [51] that the ages of acquiring these procedures are consistent with the ages of recognizing the features in the human development. However, recently Jouen and Molina [65] showed that human can manually identify the textures of objects from the neonatal stage much earlier than the age Buschnell and Boudreau had expected by recording the grasping pressure of infants on objects with different texture. This result indicates the difficulty of understanding haptics from subjective observation. Introducing measuring devices to the experiment such as the case in [65] enables objective measurement of haptics, but fails to observe the natural process.

In this section, we introduce a robot experiment which reproduces the manual touch of early infancy, squeezing and tapping, with a skin covered robotic hand. By revealing the informations extracted from the early manual touch, we try to shed light on the role of such early haptic perception in object categorization. The rest of the section is organized as follows. First we explain the robot setup, task, and exploratory behaviors applied. Then, experimental results are given followed by conclusions and discussions.

3.2.1 Robot setup

Our robotic platform can be seen in Fig. 3.1. The tendon driven robot hand is partly built from elastic, flexible and deformable materials (see [66]). The hand applies an adjustable power mechanism developed by [67]. The robotic hand has 18 degrees of freedom (DOF) that are driven by 13 servomotors and has been equipped with three types of sensors: flex/bend, angle sensors, and haptic sensors.

Bending and angle sensors

For the flex/bend sensor, the bending angle is proportional to its resistance and responds to a physical range between straight and a 90 degree bend, they are placed on every finger as

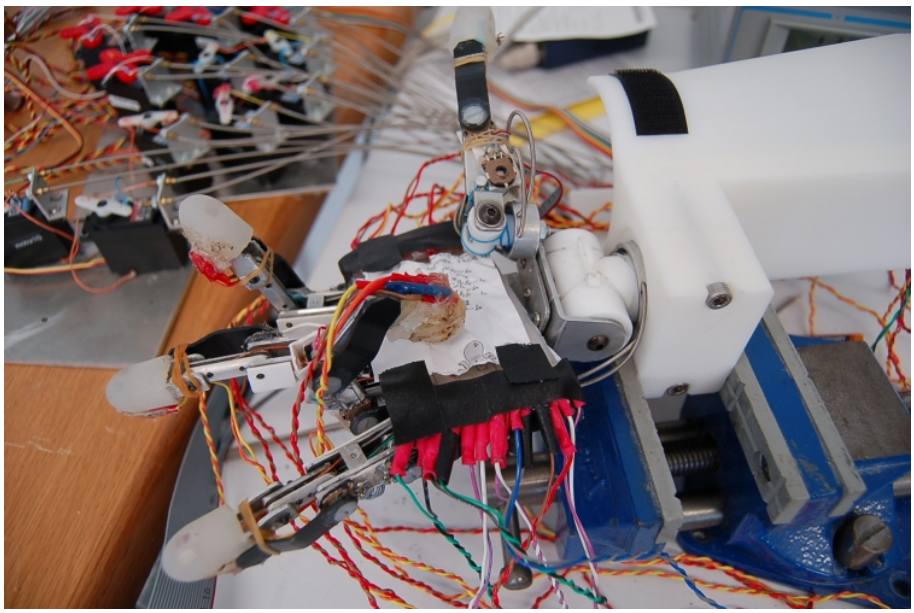
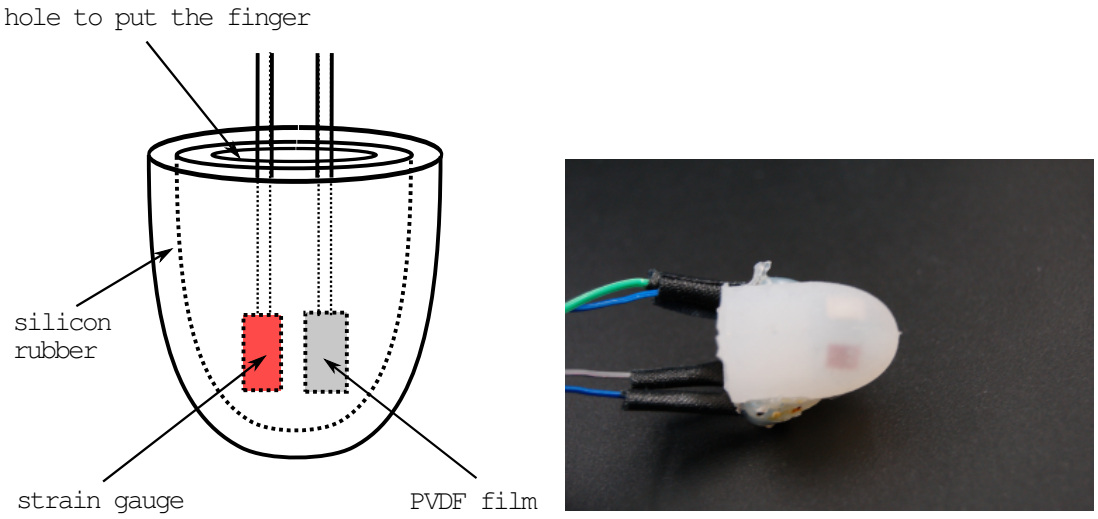


Figure 3.1: Tendon driven robotic hand. The hand is equipped with artificial skin with strain gauges and PVDF (polyvinylidene fluoride) film sensors mounted on the fingertips and the palm. The hand is exploring a piece of paper.

position sensors. Angle sensors in all the joints are provided by potentiometers.

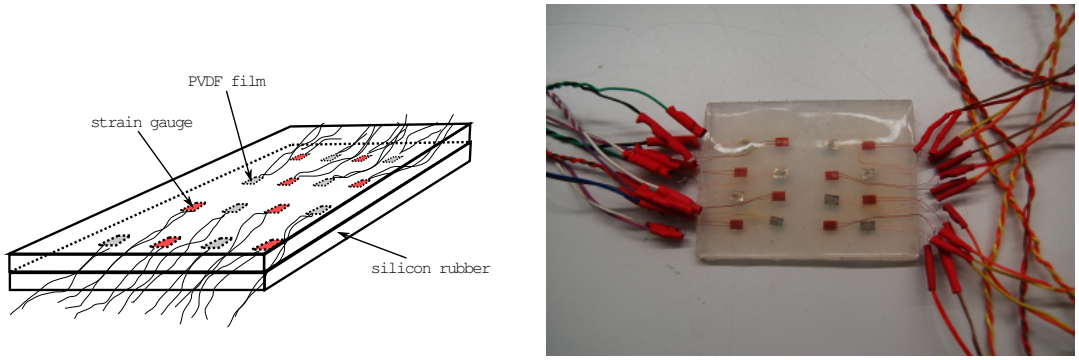
Haptic sensors

In the experiment we utilize a simplified version of the haptic sensor developed by Hosoda et al. [68]. Haptic sensors are based on strain gauges and PVDF (polyvinylidene fluoride) films sensors. The haptic sensors are located in the palm and in the fingertips of the hand. The artificial skin is made by putting strain gauges and PVDF (polyvinylidene fluoride) films between two layers of silicon rubber. The strain gauges detect the strain and works in a similar way as the Merkel cells in the human skin, whereas the PVDF films detect the velocity of the strain and corresponds to the Meissner corpuscles (MCs, see [69]) in the human skin. The PVDF films are expected to be more sensitive to the transient/small strain than the strain gauges. The shape of the artificial skin is modified to fit to the robotic hand. Sketches and photographes of the artificial skins for the fingers and the palm are shown in Figs. 3.2 and 3.3, respectively. In each fingertip there is one strain gauge and one PVDF film. In the palm there are eight strain gauges and eight PVDF films.



(a) Sketch (b) Photograph

FIGURE 3.2: Artificial skin for the fingers.



(a) Sketch (b) Photograph

FIGURE 3.3: Artificial skin for the palm.

Robot control

We control the robot hand using a TITech^{TM} SH2 controller. The controller produces up to 16 PWM (pulse with modulation) signals for the servomotors and acquire the values from the bending and angle sensors. The motor controller receives the commands through an USB port. Sensor signals from the strain gauges and the PVDF films are amplified and fed to a host computer via a CONTEC^{TM} data acquisition card at a rate of 1.6 KHz.

3.2.2 Robot task

The robot performs two exploratory procedures with the ring finger, namely: squeezing, and tapping over seven different objects of different material properties as well as the no-object condition, each object was taped on the palm of the robot hand and explored during one minute. The objects can be seen in Fig. 3.4.

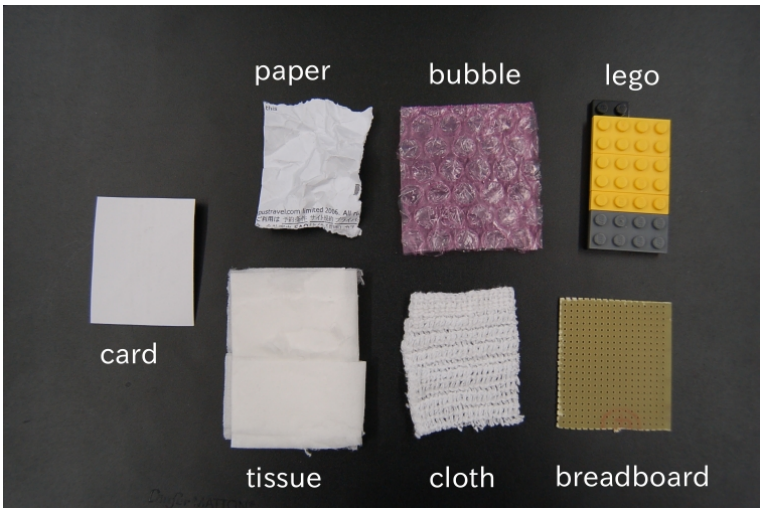


Figure 3.4: Objects with different material properties.

3.2.3 Exploratory procedures

The robotic hand actively explores different objects using the exploratory procedures: squeezing and tapping shown in Figs. 3.5 and 3.6.

FIGURE 3.5: Schematic of squeezing behavior

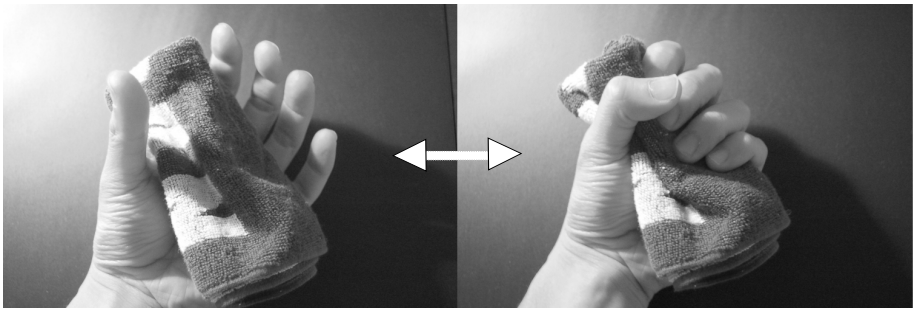


FIGURE 3.6: Schematic of tapping behavior

Squeezing

For the squeezing exploratory procedure, we drove both motors controlling the ring finger to the maximum angular position, thus making the finger to close over the palm as much as possible and squeezing the object, as described in (3.1).

$$ang_i(t) = maxAng_i \tag{3.1}$$

Where:

- ang_i is the target angular positions of the i-th finger joint (ang_L and ang_U)
- $maxAng_i$ is the maximum angular position of the i-th finger joint.

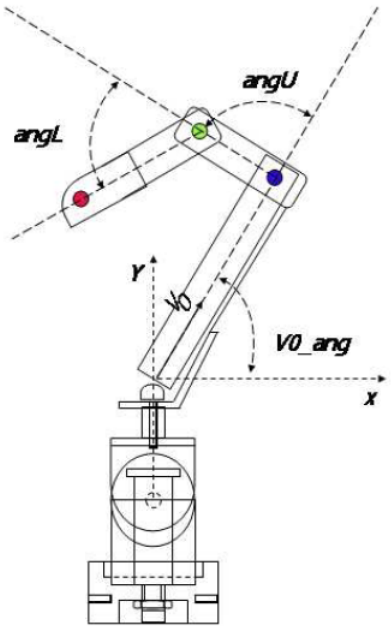


Figure 3.7: Schematic of a finger. Positions of the fingertip (red marker), the middle hinge (green marker) and the base (blue marker).

Tapping

The tapping exploratory procedure was achieved by a sinusoidal position control of the ring finger that can be described as follows:

$$ang_i(t) = A_i sin(\omega t + \phi) + B_i \tag{3.2}$$

Where:

- ang_i is the target angular positions of the i-th finger joint (ang_L and ang_U).
- A_i is the amplitude of the oscillation for the i-th finger joint.
- B_i is the set point of the oscillation (i.e. 60 degrees) for the i-th finger joint.
- ω is the frequency of the oscillation.
- ϕ is the phase delay between the oscillation

Increasing and decreasing the position of the servo motors produced the pulling of the tendons, which made the fingers move back and forth, tapping the object over the palm. Fig. 3.8 shows the result of the motion of the finger during the no-object condition.

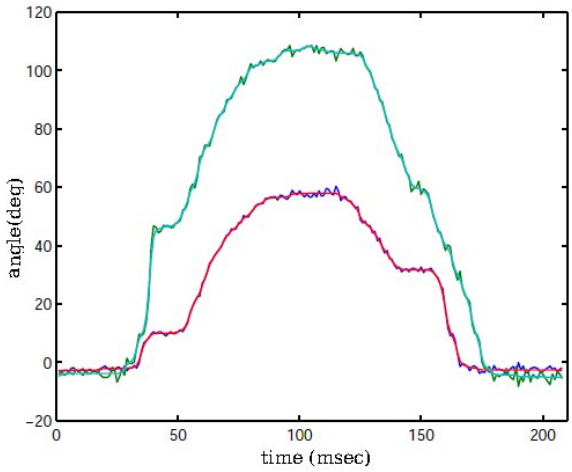


FIGURE 3.8: Kinematics of the robotic hand. sinusoidal position control. Upper plot is the angle between the middle hinge and the finger tip (angL), whereas the lower plot corresponds to the angle between the base of the finger and the middle hinge (angU).

3.2.4 Results

In the squeezing experiment, discriminative sensory values were observed from the strain gauges in the palm. Fig. 3.9 shows the typical sensory sequence of 10 second squeezing obtained from the strain gauges in the palm. 4 strain gauges on the palm were active. The yellow lines represent the output of the strain gauges during the squeezing of a piece of tissue (soft material), whereas the light green lines represent the output of the strain gauges while squeezing a circuit breadboard (hard material). A self organizing map [1] was used to observe the differences of sensory values within and between the objects. Fig. 3.10 shows the result for the squeezing exploratory procedure, the input for the SOM was the average value of the four strain gauges in the palm and the size of the SOM was 8x8. We could observe that the

[1]The self-organizing map reduces the dimension while keeping the phase relation. We used the software package SOM PAK version 3.1 [70]. The topology was a hexagonal lattice, the neighboring function type used was bubble.

objects with stiff surface comes in the middle while the objects with soft surface comes in the sides.

In the tapping experiment, discriminative sensory values were obtained from the PVDF film on the fingertip. Fig. 3.11 shows a typical sensory sequence of a tapping experiment during 2 sec from the PVDF film on the fingertip of the ring finger. The color correspondence is as follows: no-object condition(red), breadboard(light green), card(blue), lego(light blue), paper(pink), tissue(yellow), cloth(black), bubble(white). The larger output corresponds to the moment when the finger taps over the object and the smaller output corresponds to the moment when the finger is pull back and leaves the object. Fig. 3.12 shows an analysis with a self organizing map for the tapping experiment. A self organizing map with same structure but different size (16x16) was utilized for the analysis. The input in this case was the values from the PVDF films on the fingertip. We can observe that soft surface materials are on the left/upper part whereas hard surface objects are in the middle. The case of no objects were also different from the other. This result was caused by a bumping behavior of the finger on the palm by the elasticity of the skin.

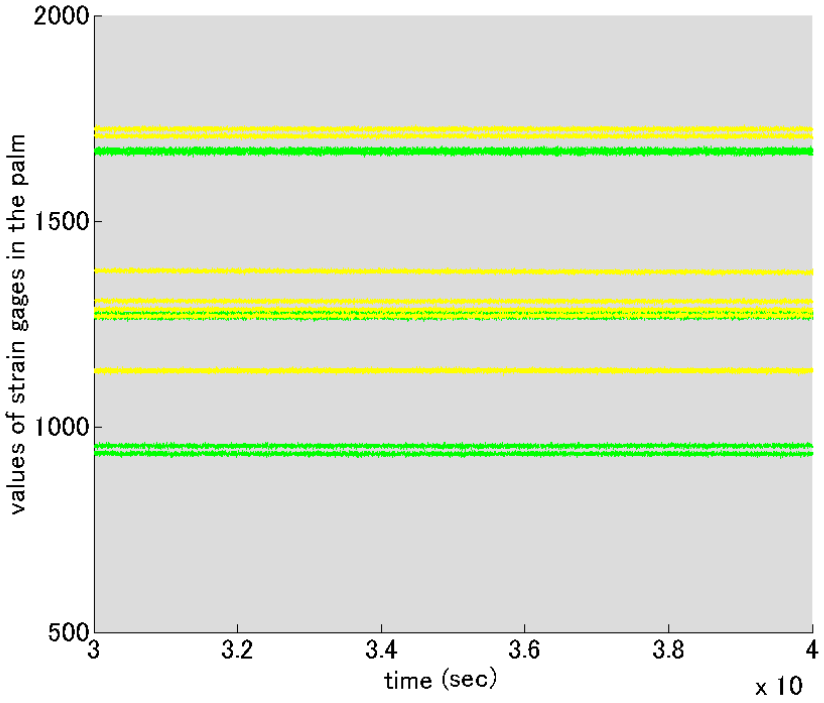


FIGURE 3.9: Squeezing exploratory procedure. Output of the four strain gauges located on the palm of the robotic hand while exploring a piece of tissue(yellow) and a circuit breadboard (light green) during 10 sec.

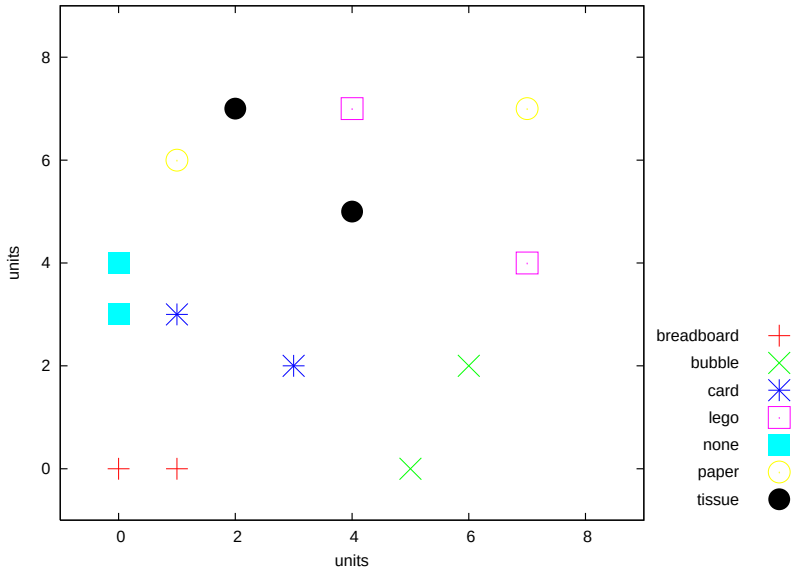


FIGURE 3.10: Classification of seven objects plus the no-object condition by a SOM using the "Squeezing" exploratory procedure.

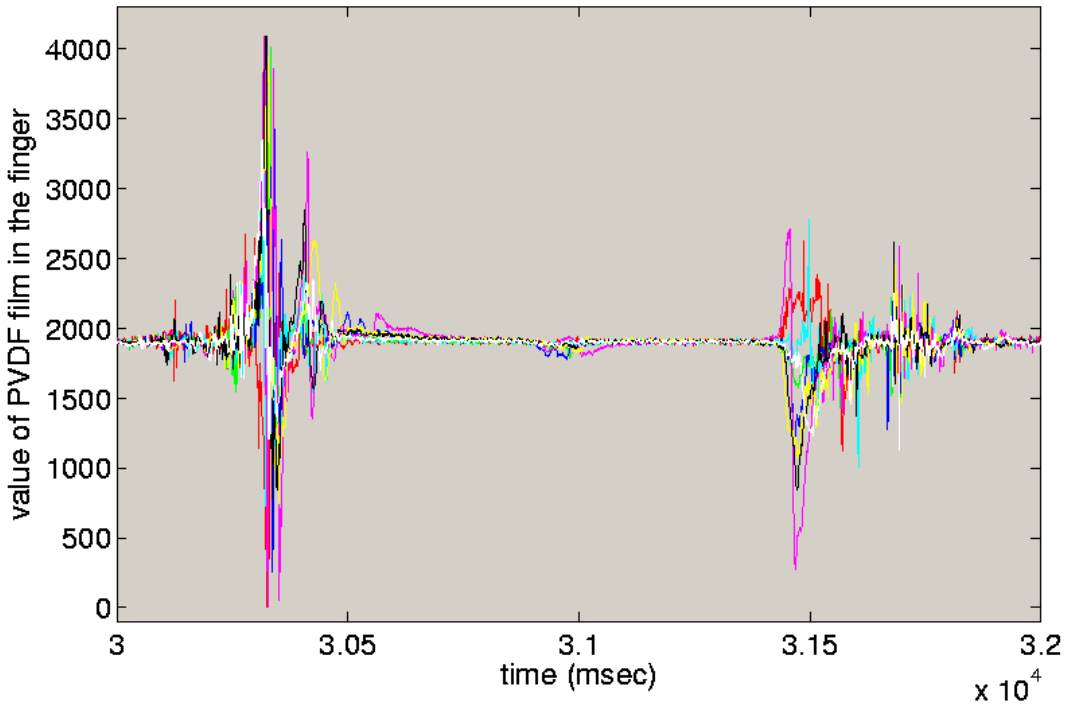


Figure 3.11: Tapping exploratory procedure. Output of the PVDF film located on the fingertip of the ring finger while exploring the objects in Fig. 4.8 during 2 sec.

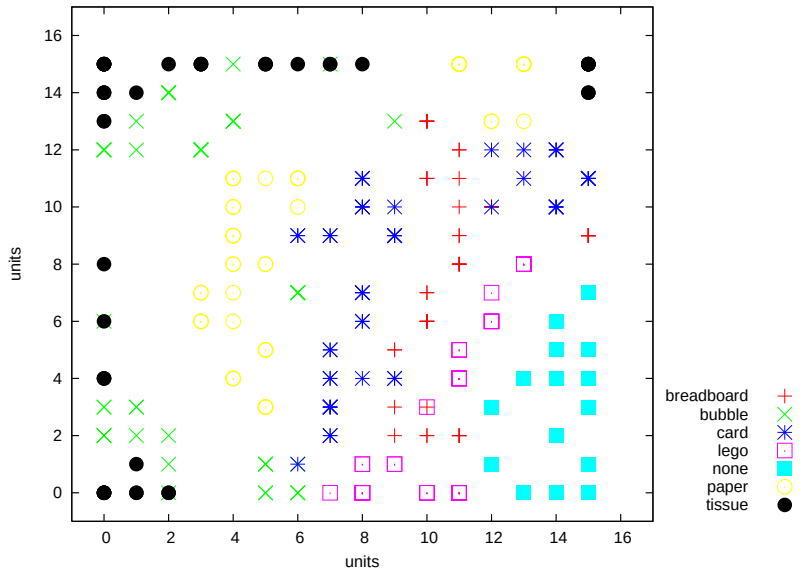


Figure 3.12: Classification of seven objects plus the no-object condition by a SOM using the "Tapping" exploratory procedure.

3.2.5 Conclusion

The mechanism behind the result of the squeezing experiment can be explained as follows: When the object is soft, the pressure of the fingertip on the palm is perpendicular, whereas, when the object is stiff, the pressure on the palm depends on the edge of the object. On the otherhand, the results of the tapping experiment should be formed from the dynamic deformation of the fingertip when it hits and leaves the object. If the object is soft, there should be a moment when the finger goes into the object producing relatively gentle change in the pressure. If the object is stiff, the pressure rises rapidly. If the object is sticky, the object sometimes follows the finger when it leaves. To conclude, both early exploratory manual touch, squeezing and tapping, are capable of recognizing the stiffness of the objects through the static and dynamic deformation of the manual skin. Furthermore, we could expect that squeezing behavior tells whether there are edges on the object, and tapping behavior detects the stickiness of the surface. Since such surface state of the object are important to identify categories, we could infer that the early manual touch of infants and neonates could play a role in object categorization.

3.3 Dynamic touch experiment

Several attempts are made to implement the exploratory behavior of infants to robots to obtain multimodal categories of objects [53] [54] [55] [56]. However, most work conducted till today implemented static behaviors for exploration. The problem of static behaviors such as grasping and hitting can only obtain information of object parts close to the contact point. Therefore, existing approaches fail in recognizing the object category when the size, shape, or contact condition changes. It is easy to imagine how much this problem limits the performance. In order to avoid such a problem, we focus on another frequently observed behavior of infants, shaking. Although the difficulty of measuring and controlling the behavior have kept it from being a hot topic in the field of developmental psychology, there are enough convincing reasons to do so. It is pointed out by Turvey [71] and his colleagues that shaking behavior gives rich information of the whole object. This effect eases the acquisition of object categories which can be generalized to objects with different sizes, shapes, and contact conditions. The rhythmic actuation in shaking behaviors also realizes entrainment [72] which enables stable recognition under rough control.

In this section, we introduce results of a robotic experiment which investigates the effectiveness of shaking behaviors in object category acquisition. Although several existing studies are found which shows that shaking behavior helps object recognition of rigid objects by detecting the momentum of inertia [57], we show for the first time that shaking behaviors are also effective for acquiring humanlike daily object categories by introducing the auditory information processing of the cochlea. The rest of the section is organized as follows. First, we introduce some related work and describe our system of categorization. Next, we explain the experiment design and show the results with some analysis. Finally, conclusion and discussion are given.

3.3.1 Related work

The role of shaking behavior in exploring object properties was initially pointed out by the research group of University of Connecticut headed by Turvey [73]. They gave experimental results which imply that shaking behavior, also referred to as dynamic touch, gives information of object lengths, shapes, and contact conditions. Their work still remains as one of the largest efforts on this topic. However, the work dealt with only identification of rigid objects such as rods with different length. In the field of developmental psychology, very few studies on shaking behavior are found [74] [75]. The lack of such studies comes from the difficulty of objective measurements. Several studies based on synthetic approach are found for shaking where Suzuki et al. [76] developed a humanoid which discriminates two cylinders with different length by detecting the momentum of inertia, Williams et al. [77] realizes generation of robot arm movements which exploit the dynamics of the objects with neural oscillators, and Nabeshima et al. [78] showed how shaking can be utilized for detecting affordances of tools through simulations. The idea to utilize frequency responses is also introduced by Fregolent and Sestieri [79] to detect inertia properties of objects. However, the possible roles of shaking behavior in object categorization are not well investigated.

3.3.2 System

We propose a system of acquiring object categories from sensory sequence obtained through shaking behavior. We focus on auditory sense in this system based on the finding in psychological study that tells us that infants shake objects more when they produce sound, and we implemented the system using the auditory sensing devices that have high sensitivity and quick response with low cost. Fig. 3.13 shows an overview of the system. First, a rhythmic actuation on the arm produces a stable cyclic behavior under rough control by virtue of entrainment. The actuation can be of any kind as long as it is cyclic; simple sine curves, square waves, actuations with neural oscillators, etc. Then, the obtained auditory sensory sequences are subsequently processed by a Fourier transform circuit to obtain amplitude spectrums of them. Such circuits are also found in human ears known as the cochlea [80]. Finally, the amplitude spectrums are utilized as feature vectors to form object categories

by a clustering method. When the object is shaken, distinctive oscillations depending on the category generate at some parts of the object and propagate through air to reach the microphone. The phases of the oscillations on the microphone will vary by the distance of the microphone and the oscillating part of the object. Consequently, the extraction of the amplitude spectrum removes the information of size, shape, amount, and contact conditions, and eases robust object categorization.

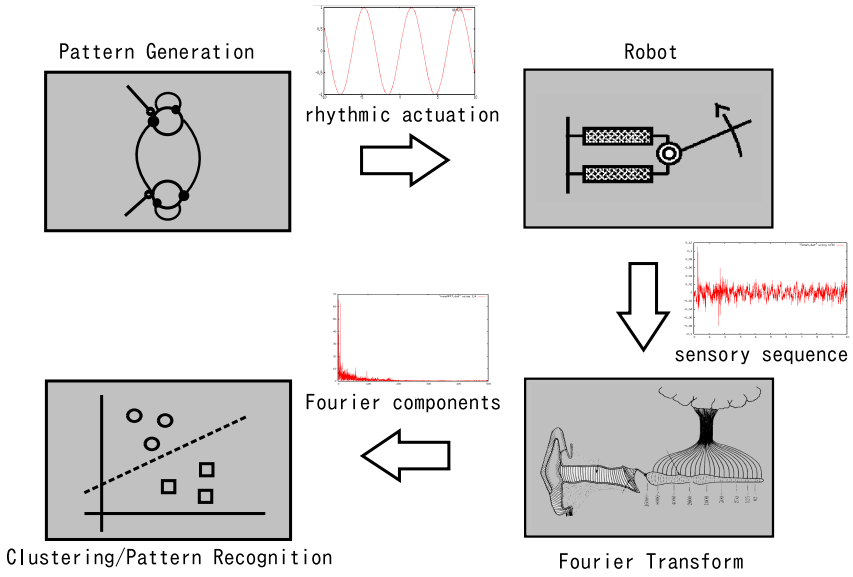
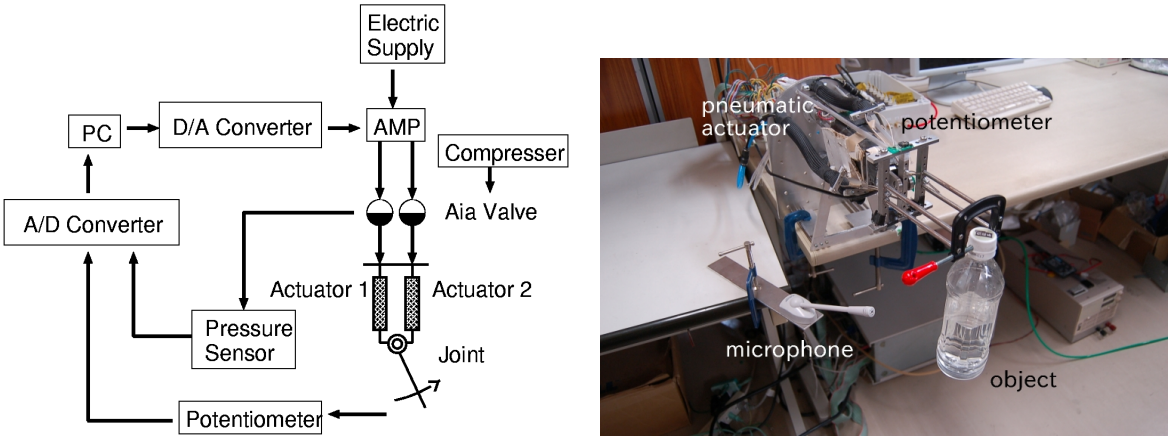


Figure 3.13: An overview of the proposed approach

3.3.3 Experiment Design

In the experiment, we utilize a robotic arm with McKibben pneumatic actuators. The robot is equipped with a microphone to obtain the auditory data and a potentiometer to obtain joint angle data. Fig. 3.14 shows a photograph of the robot with a graph of the control system. The arm shakes the objects in the horizontal plane for simplicity reducing the effect of gravity on the shaking behavior. A stable limit cycle behavior was realized by putting air in an anti-phased manner to the antagonistic actuators controlling the valve gates. The duration of shaking was approximately 10 seconds. We utilize self-organizing maps [81] to



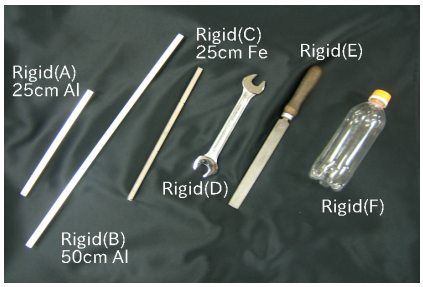
(a) Control system (b) Photograph

FIGURE 3.14: Robot used in the experiment

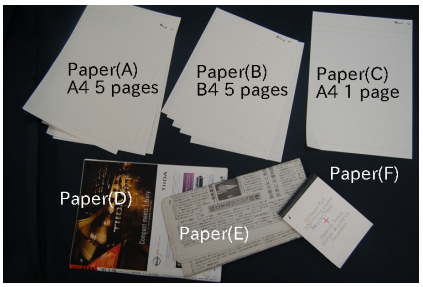
examine whether the feature vector is suitable for acquiring the categories. The distance measure d for the self-organizing map is given as follows where p_{1i} and p_{2i} are components of the feature vectors and N is the number of these components.

$$d(\boldsymbol{p}_1, \boldsymbol{p}_2) = \sum_{i=1}^{N} |p_{1i} - p_{2i}|$$

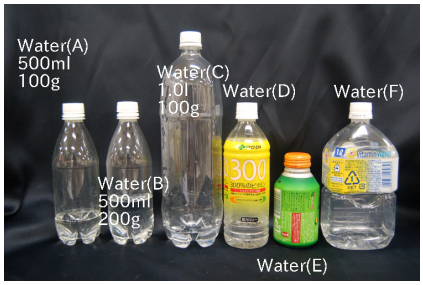
The task was to acquire and recognize three object categories, namely rigid objects, paper materials, and bottles with water. The objects utilized in the experiment are shown in Fig. 3.15. The first few objects have controlled variance, which means that only one property such as size or amount is different. The latter three objects have uncontrolled variance. Since daily object categories allow so many properties to change, objects with such daily variety should be used to evaluate the ability of category acquisition. The rigid objects differ in size and shape, papers differ in size, shape, thickness, page numbers, and bottles differ in size, shape, amount of water inside, and material of the bottle. Since the objects have different appearances including transparent cases, it would be difficult to categorize them with only the visual information due to the changes in light condition and backgrounds. Categorization with grasping would also fail since the sense would vary by contact conditions.



(a) rigid body objects



(b) paper materials



(c) bottles with water

Figure 3.15: Objects used in the experiment

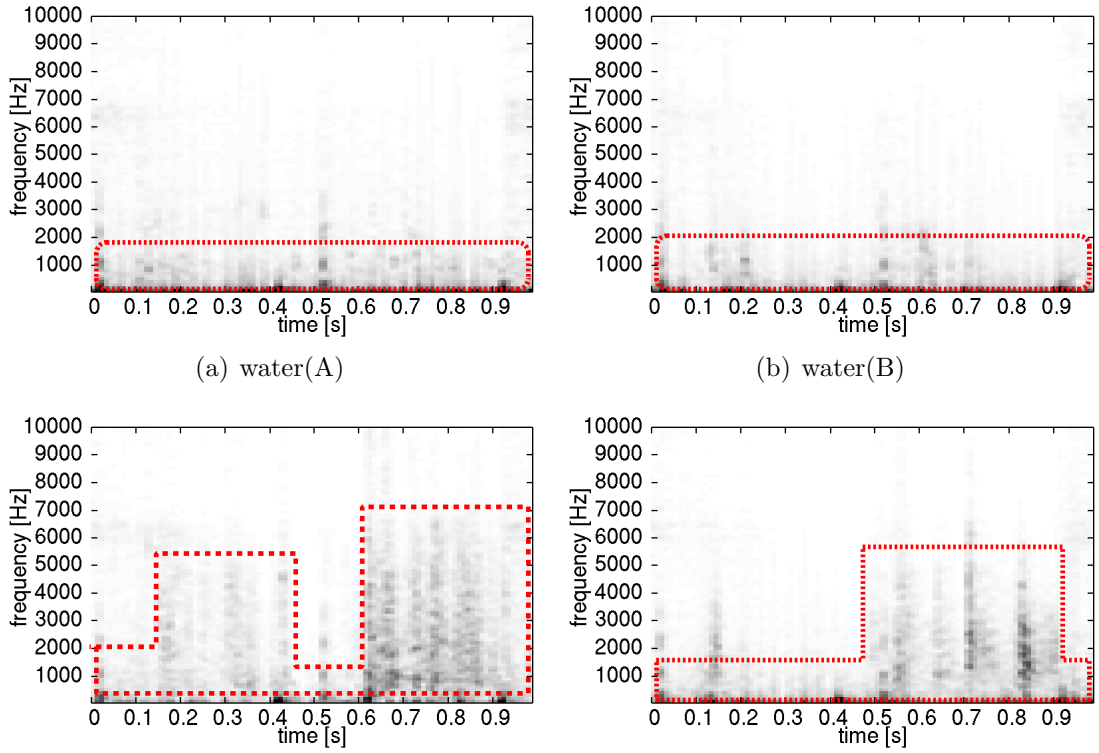
In order to show that the category recognition is also independent to the grasping posture and shaking frequency, two additional conditions are included for each category. The conditions for all the trials in the experiment are shown in the Appendix. Shaking for each condition was repeated for 5 trials.

3.3.4 Results

First, the spectrograms, the time sequences of amplitude spectrum, are shown to examine the characteristics of the feature vector. Then, analysis with self-organizing maps is given to investigate whether the amplitude spectrum is a suitable feature vector for object categorization. Pattern recognition performance with leave-one-out method is also included as a quantative measure of the suitability.

Spectrograms

Representative spectrograms for bottles of water and paper materials are given in Fig. 3.16. The spectrogram of rigid objects were flat since rigid objects do not produce much sound. The spectrograms were qualitatively similar within the categories and qualitatively different between the categories; spectrograms of bottles of water have relatively common amplitude in low frequency regions, whereas paper materials have amplitudes partially reaching relatively higher frequency regions.



(a) water(A)

(b) water(B)

(c) paper(A)

(d) paper(B)

FIGURE 3.16: Spectrogram of sound generated by shaking.

Analysis with a self-organizing map

In order to visualize the differences of the amplitude spectrums within/between the object categories, we fed it into self-organizing maps [81]. The self-organizing map reduces the dimension while keeping the phase relation. The results of the clusterings for feature vector of raw auditory data and its amplitude spectrum is shown in Figs. 3.17 and 3.18, respectively. Both inputs had $N = 200$ components; the raw auditory data was a time sequence of sound strength of 200 discrete points in a single shaking cycle, whereas amplitude spectrum was a time average of the amplitude spectrum for 100Hz-20KHz also digitized into 200 discrete values. Amplitude spectrums are calculated from the spectrogram by taking the time average. Both networks of self-organizing map had two layers with 32 units each and the best matching unit for each trial is plotted on the 32×32 grids. Circles represent bottles of water, holizontal squares represent paper materials and inclined squares represent rigid objects. The figures show that clustering of object categories fails with raw auditory data, but is successful with the amplitude spectrum of the auditory data. The representational points for the objects with varying size, shape, amount, contact conditions, shaking frequency, and so on, are found together according to the category with the amplitude spectrum. Several objects clustered near the boarders are newspapers, magazines, and Alminium cans, which produce relatively different sounds from the others. Further information processing are recommended to improve the robustness of the categorization. Pattern recognition performance with leave-one-out method is also given in Table 3.1 as a quantative measure of the suitability of the feature vector.

Table 3.1: Leave-one-out recognition rate with the feature vectors.

	water (%)	paper (%)	rigid (%)	all (%)
raw auditory data	40.0	25.0	82.5	49.2
amplitude spectrum	92.5	95.0	100.0	95.8

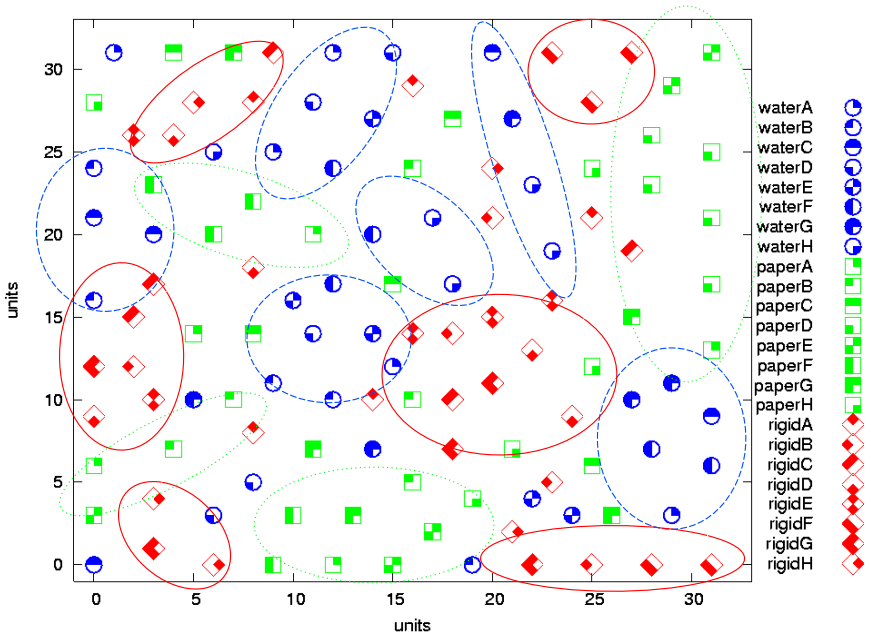


FIGURE 3.17: SOM analysis result with raw auditory data

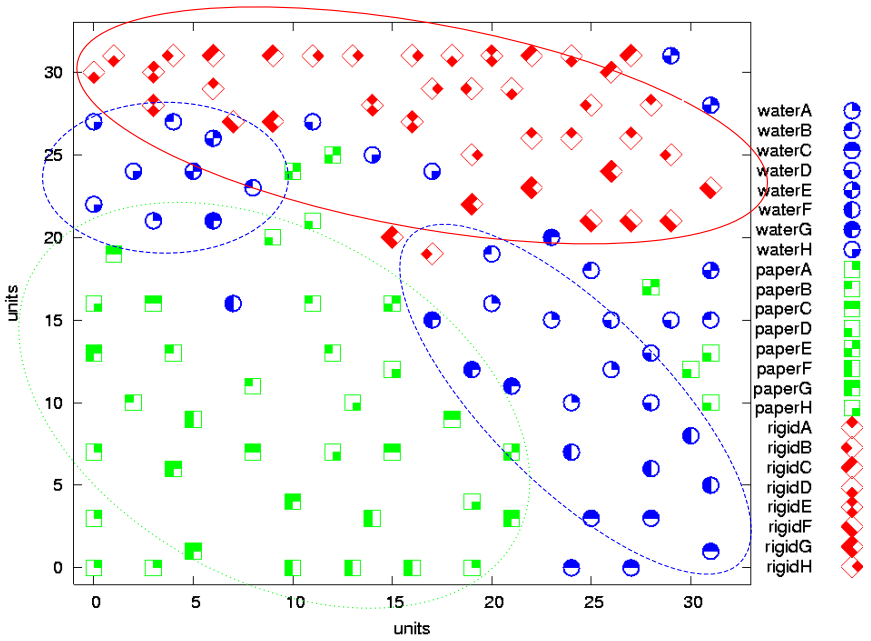


FIGURE 3.18: SOM analysis result with amplitude spectrum of auditory data

3.3.5 Conclusion

I considered active auditory sensing with dynamic touch like shaking as one possible method to obtain such invariant feature. By obtaining the amplitude spectrum of auditory data during shaking, the agent can detect distinctive oscillations which specify the category. This idea was supported by the result of the experiment which showed that the auditory amplitude spectrums obtained by shaking paper materials, bottles of water, and rigid objects were similar within the categories and differed between the categories even though the objects within each category varied in size, shape, amount, contact conditions, and so on. The idea that amplitude spectrums of auditory signals include the information for object categorization could be estimated from our daily experience, but the fact that the extraction of the amplitude spectrum plays an essential for the categorization was shown for the first time in this experiment. The spectrogram of auditory data from the trials were qualitatively similar within the categories but qualitatively different between the categories. This indicates that the amplitude spectrum obtained from different categories can be distinguished and the performance of object recognition will improve by having more experience. Shaking the object with multiple different contact conditions or multiple different actuations are possible methods to improve the robustness. Since the method extracts invariant feature of the object independent to the distance to the object or the frequency of the shaking motion, it may explain how a robot and a human which have very different body structure could obtain similar categories, although it is just one possible approach to the issue.

Chapter 4

Lexicon acquisition based on behavior learning

4.1 Introduction

Human infants could specify word meanings from diverse possiblities with very few teaching [13]. The question how infant manage to learn the lexicon in such a manner, the Gavagaii problem [14], have led a long lasting debate [15] [16]. From the recent findings suggesting that words directly grounded in physical embodiment form the basis for acquiring more abstract words and syntax [82], the issue of symbol grounding [83] is now considered even more fundamental. Several recent linguistic experiments on word generalization indicate the role of behaviors in object categorization. Nelson et al. [17] showed that even 2-year-old children can associate names to new objects in accordance with objects' functions when the function of the object is apparent to them. On the other hand, Kobayashi [18] showed that an caregiver's actions on objects influence children's inferences about word meanings. These observations suggest the possibility that equivalence of behaviors afforded to an agent, the affordance [25] of the objects, could be the key for object categorization. The idea seems feasible from pragmatic perspective since such word meanings relates to the function of the objects. When a person asks for a chair, it could be any object that affords the behavior of sitting on it.

Inspite of the attractive and natural idea, current robot systems are not capable of such lexicon acquisition. Most robot systems capable of acquiring object representation from physical experience [53] [54] [55] [56] [58] [59] are equipped with fixed behaviors and thus not capable of acquiring various behaviors afforded by the object. The work on symbol grounding by Sugita and Tani [84] showed how compositional syntax could emerge from an attempt to acquire generalized correspondence between word sequences and sensorimotor flows. This work shows how word learning by behaviors is beneficial for extending to language acquisition. However, here again, the behaviors for the objects are mostly given by the designer.

This chapter introduces the development of robotic system capable of object categorization and lexicon learning based on learned equivalence of behaviors afforded by the objects. In order to acquire the word "cylinder" for example, the learner acquires the behavior to face the lateral aspect of the object and roll it. The behaviors with the objects are learned and identified by a multi-module learning system [85] modified for reinforcement learning. We show a method to associate sensorimotor concepts represented in the multi-module learning system with labels without assuming the labels to be given simultaneously with the activation of those learning modules. The chapter is organized as follows. First, we explain our approach of lexicon acquisition based on behavior learning. Then, experimental results show the validity of our approach. Finally, conclusions and future issues are given.

4.2 Lexicon Acquisition based on Behavior Learning

4.2.1 Basic Idea

In our study, we address the task of learning labels for objects given by a caregiver (Figure 4.1) and associating appropriate labels to new objects based on similarities in functions or in affordances. The learner is considered successful when it answers appropriate labels for given objects in view. The actual lexicon acquisition task involves various difficult tasks such as extracting individual words from continuous utterances and attending to the object intended by the caregiver. However, in order to focus on the label association issue, we do

not deal with these tasks here.

The lexicon acquisition process of the system consists of four sub-processes shown in the box bellow.

A Learn object-oriented behaviors.
B Categorize (photometric) visual feature space based on object-oriented behaviors.
C Categorize (photometric) visual feature space based on labels given by the caregiver.
D Learn the correspondence between object-oriented behaviors and labels.

A is an process where the learner learns the behavior for given objects. The behaviors specified for certain affordance or function of objects such as cutting behavior for the label "scissors" is referred to as object-oriented behaviors. During this behavior learning process, the learner builds models of object behavior caused by the robots behavior which can later be utilized to identify behaviors that the objects afford. **B** is a process of categorizing the visual feature space based on the object behavior model acquired in process **A**. This process enables the learner to identify the behavior for the object in view without physical interaction. As for visual features, we adopt photometric features such as color histograms or edge histograms. Although visual categories related to object-oriented behavior is acquired by **A** and **B**, the learner still needs to associate labels to the object-oriented behaviors in order to talk about it. This association is not a easy task since categories of behaviors are gradually obtained and not all labels are related to those categories. In order to solve the association task, the learner categorizes the visual feature space based on the labels (**C**), and learn the association by matching the categories of behaviors and of labels (**D**).

4.2.2 System Overview

The overview of our system for lexicon acquisition based on behavior learning is shown in Figure 4.2. Inputs to the system are classified into three types: namely, *photometric features* for object identification, *state variables* for controlling the objects such as current position and orientation of objects, and *labels* given from the caregiver. The idea to separate visual

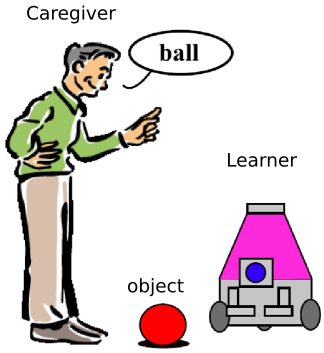


FIGURE 4.1: Environment of lexicon acquisition.

information of objects into two main categories, photometric features and state variables, comes from the study by Milner and Goodale [86] which indicates that visual information for identification and control are handled in different processes in human brains. Similar separation of visual information is also adopted in the system of Fitzpatrick and Metta [87] to learn the affordance of objects. As previously mentioned, process **A** is realized by a *multi-module reinforcement learning system* taking state variables as input. Each module of the multi-module reinforcement learning system corresponds to a particular object-oriented behavior. As **A** performs, **B** easily performs by categorizing the photometric feature space according to the identification of object-oriented behaviors. On the other hand, the system also categorizes the photometric feature space according to the labels given from the caregiver (**C**). This means that two different categorization is performed on the same photometric feature space. *Adaptive networks* are adopted for these categorizations. Finally, process **D** is realized by learning the correspondence between object-oriented behaviors and labels by a *Hebbian network*, connecting a behavior and a label whose category is selected for same photometric features. Details on each learning system are explained in the following sections. Although the different learning processes are explained one at a time, the system is designed to run all the learning processes in parallel. Scheduling of lexicon acquisition is introduced only for the simplicity of showing the results.

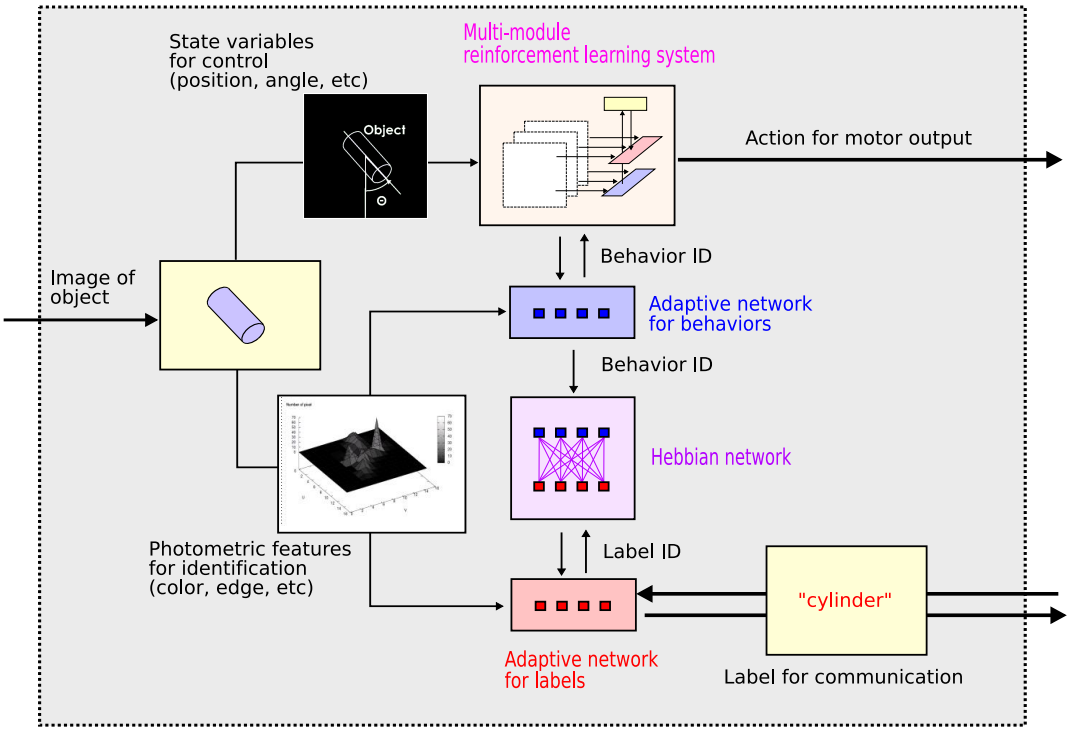


Figure 4.2: Sketch of system for acquiring lexicon based on behavior learning.

4.2.3 Learning and identifying object-oriented behaviors

We adopt the multi-module reinforcement learning system shown in Figure 4.3 for learning and identifying object-oriented behaviors. The system consists of multiple learning modules each of which consists of a predictor and a planner, and a gate to select the appropriate module based on reliability representing the accuracy of goal-directed state transition prediction by each learning module. In the current system, one-to-one correspondence between learning modules and object-oriented behaviors is assumed. The system learns object-oriented behaviors as state-action mappings, and identifies the object-oriented behaviors based on the reliabilities of the learning modules. If no learning module with sufficient reliability is found, a new learning module is assigned to learn a new object-oriented behavior. We chose Q-learning [88] associated with *state transition models* and *reward prediction models* as the reinforcement learning method. The method can acquire the behaviors with relatively little prior knowledge on the task or the environment.

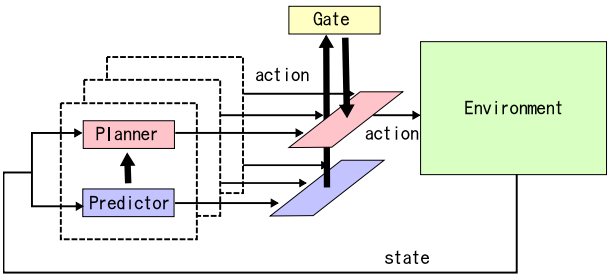


FIGURE 4.3: Multi-module reinforcement learning system.

Behavior learning

In general reinforcement learning, interaction between learner and environment is modeled as shown in Figure 4.4. In every time step, learner obtains a discrete representation of the current state $s_t \in \mathcal{S}$ (S is the set of possible states), and selects an action $a_t \in \mathcal{A}(s_t)$ ($\mathcal{A}(s_t)$ is the set of possible action at state s_t). Then the next state $s_{t+1} \in \mathcal{S}$ and reward $r_{t+1} \in \mathcal{R}$ is determined, depending only to the state and action selected by the learner. Task of reinforcement learning is to choose a policy $a = f(s)$ which maximizes the decaying sum of reward shown below,

$$\sum_{n=0}^{\infty} \gamma^n r_{t+n}, \tag{4.1}$$

where γ is the decay factor ($0 < \gamma < 1$).

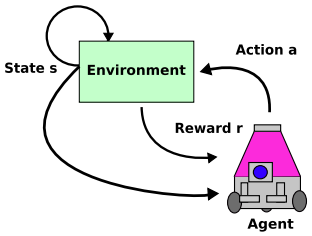


FIGURE 4.4: Basic model of learner-environment interaction in reinforcement learning.

The predictor in each learning module builds two models of the environment through the interactions with the objects. One is a *state transition model* which is a set of probabilities of all state transitions.

$$\hat{\mathcal{P}}^a_{ss'} = Pr\{s_{t+1} = s' | s_t = s, a_t = a\} \tag{4.2}$$

Another is a *reward prediction model* which is a set of expected reward values for all state-action sets.

$$\hat{\mathcal{R}}_s^a = \sum_{s'} E\{r_{t+1}|s_t = s, a_t = a\} \tag{4.3}$$

As the *state transition model* and *reward prediction model* are built, the planner for each learning module calculates the *action value* $Q(s, a)$ (a set of expected decaying reward sum for every state-action set) by simulating the learning process offline. The offline learning process takes place at the end of each trial of interaction with the environment. All the state transition data from a continuing interaction are assumed to be coming from the same object.

$$Q(s, a) = \sum_{s'} \hat{\mathcal{P}}_{ss'}^a [\hat{\mathcal{R}}_s^a + \gamma \max_{a'} Q(s', a')] \tag{4.4}$$

When $Q(s, a)$ converges, the rational policy is given as follows.

$$f(s) = arg \max_{a \in \mathcal{A}(s)} Q(s, a) \tag{4.5}$$

In order to collect data for building the models, the learner selects actions for the objects by first specifying the best matching learning module and then selecting the action within the learning module. A learning module is usually selected based on its reliability defined in the next Section. However, in the initial phase of interaction where no state transition is observed, the learner selects the learning module corresponding to the best matching object-oriented behavior for the given photometric feature of the object. The adaptive network is used for this selection and discussed in detail in Section 4.2.4. Once a learning module is selected, the learner selects the action based on ϵ-greedy method. For more effective identification, the learner can choose the action which gives the most different state value transition among the learning modules.

Behavior Identification

We calculated the reliability for each learning module as

$$rel = DQ_{threshold} - \Delta Q(s_t, a_t) \tag{4.6}$$

in which $\Delta Q(s_t, a_t)$ is the action value error defined as

$$\Delta Q(s_t, a_t) = \left| r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \right|. \tag{4.7}$$

This means that when the learner encounters an object, it identifies the object-oriented behavior by choosing the learning module with the most accurate prediction of the state value transition. We adopted this criterion instead of common state transition predictions in order to put more importance to task-oriented state changes. When action value errors of all existing behavior modules exceed a predefined threshold $DQ_{threshold}$, all reliabilities have negative values and a new learning module is assigned to learn the new behavior. The threshold was balanced to avoid redundant computation but to acquire enough categories for object handling.

4.2.4 Categorization of photometric feature space

We adopted adaptive networks, a modified radial basis function neural network, for the categorization of the photometric feature space. The method is more tolerant to noise than simple nearest-neighbor methods and considered as a model for categorical perception of biological systems [89]. The method is also adopted by Steels and Belpaeme [90] to acquire categories for words.

Adaptive network consists of a set of network, each of which corresponding to a single category. Network of each category outputs a scalar value $y(\boldsymbol{x})$ for a input of photometric feature vector $\boldsymbol{x}$ and the category with the largest network output is selected as the best matching category. Each category network consists of locally reactive units whose responses are greatest at a central value $\boldsymbol{m}$, and decay exponentially around this central value. Response of local unit j is

$$z_j(\boldsymbol{x}) = e^{-\frac{1}{2}(\frac{d(x, m_j)}{\sigma})^2} \tag{4.8}$$

where $d(\boldsymbol{x}, \boldsymbol{m}_j)$ is the distance between central value $\boldsymbol{m}_j$ and input vector $\boldsymbol{x}$. Constant values are set for variance σ. Output of the network for category k is

$$y_k(\boldsymbol{x}) = \sum_{j=1}^{J} w_{kj} z_j(\boldsymbol{x}) \tag{4.9}$$

where J is the number of locally reactive units. Sketch on calculation of the output for category networks is shown in Figure 4.5.

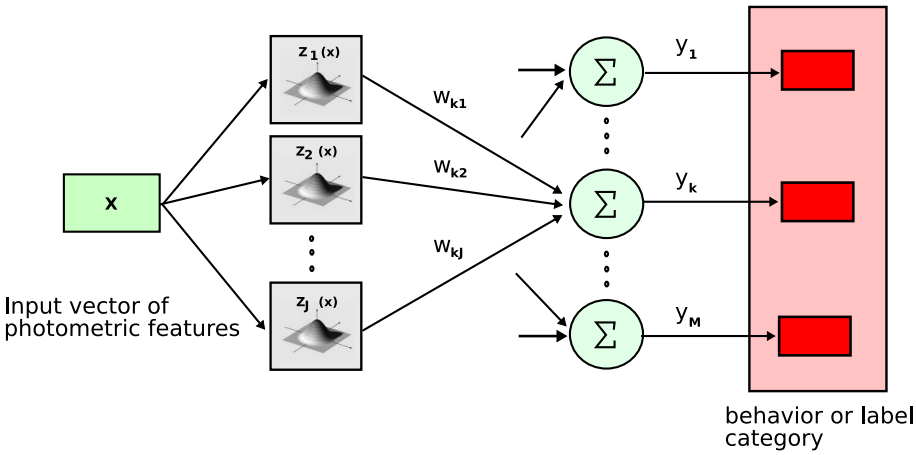


Figure 4.5: Sketch of adaptive network for behavior or label. Category with the largest network output is selected.

The system categorizes the feature space by modifying the network weights and adding new local units. When a training data set (photometric features with a label or photometric features with a behavior) is given and categorization is successful, the network weight of the matching category k is increased as shown below.

$$w_{kj} \longleftarrow w_{kj} + \beta z_j(\boldsymbol{x}) \tag{4.10}$$

If the categorization is not successful, a new local unit with a central value of the unsuccessfully categorized input is assigned and the weight to the correct category is set to a predefined value of w_0 [1]. When new categories are added, such as the case when new behavior modules are assigned or new labels are taught, new category network is assigned with local units capable of classifying inputs for those new categories. The weights of all local units of all categories are decreased whenever the network is modified.

$$w_{kj} \longleftarrow \alpha w_{kj} \ (for\ k = 1, \cdots, K,\ and\ j = 1, \cdots, J) \tag{4.11}$$

[1] Other weights are set to 0.

This process enables the system to forget unused categories, and keep adapted to changes of the environment. $\beta \in [0, 1]$ is a learning rate, and $\alpha \in [0, 1]$ is a decay factor.

4.2.5 Learning relation between object-oriented behaviors and labels

We adopt Hebbian network for learning relation between object-oriented behaviors and labels. The Hebbian network connects the nodes of object-oriented behaviors and the nodes of labels, and modifies the weight of the network so that corresponding nodes are strongly connected. Whenever the learner captures an object, the learner extracts the photometric features of the object and selects the best matching behavior and the label utilizing both adaptive networks. In case where behavior m and label n is selected, the weight between those two nodes W_{mn} is increased as follows

$$W_{mn} \longleftarrow W_{mn} + \delta_{inc}, \tag{4.12}$$

and weights of other connections to node m or n are decreased with δ_{inh} to disregard labels unrelated to object-oriented behaviors.

$$W_{st} \longleftarrow W_{st} - \delta_{inh} \quad (s = m \ or \ t = n) \tag{4.13}$$

δ_{inc} and δ_{dec} have positive values smaller than 1. When new nodes are added, the weights are initialized to 0s.

4.2.6 Label association policy

The process of selecting labels for objects in view is shown in the following. The process includes label selection based on similarity of behaviors that the object affords.

1. Extract photometric features of objects.

2-A. Select the best matching behavior utilizing the adaptive network for behaviors.

2-B. Select the best matching label utilizing the adaptive network for labels.

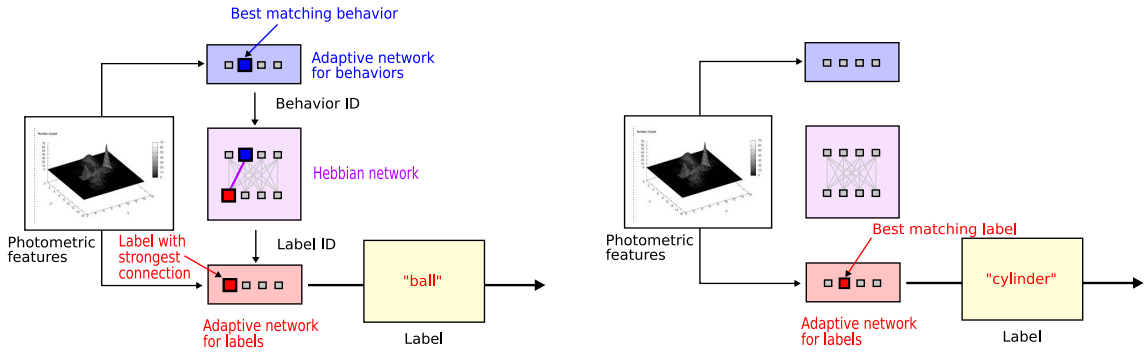
3. Compare the outputs of the selected behavior category network and the selected label category network.

4. If the output of behavior category network is larger than that of label category network and there exists a label unit in the Hebbian network which is connected to the unit of the selected behavior with a weight higher than a predefined value W_{limit}, go to step **5**. Else, go to step **6**.

5. Output the label with strongest connection to the behavior selected in step **2-A**.

6. Output the label selected in step **2-B**.

Figures 4.6 (a) and (b) show the information flow for selecting the label in step **5** and step **6**. The learner outputs labels according to behavior when it knows the behavior for the object and has acquired the label for it. Learner with low W_{limit} associates labels more often with behavior categories, whereas learner with high W_{limit} are more conservative on such association.



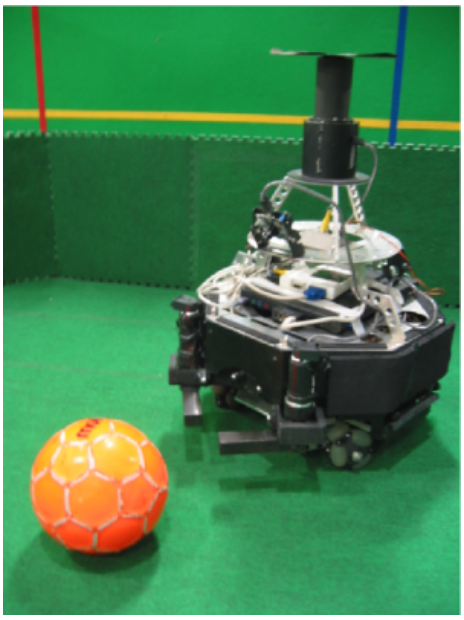
(a) label guessing by behavior category (b) label guessing by label category

FIGURE 4.6: Label guessing process.

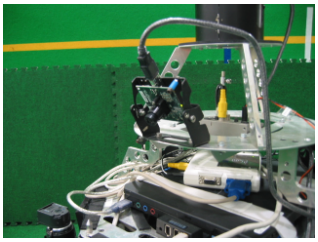
4.3 Experiments

4.3.1 Task

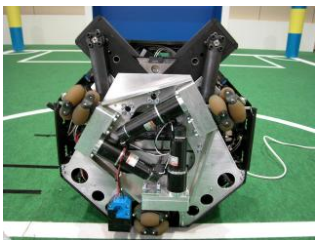
To show the validity of our system, we implemented the system to a mobile robot (Figure 4.7) who learns lexicon about objects with different rolling preferences. We used the objects shown in Figure 4.8, and gave labels namely “ball”, “box”, “cylinder”, and “car”. After presenting the objects shown in Figures 4.8 (a), (b), (c), and (d) paired with the labels corresponding to them, we introduced new objects shown in Figures 4.8 (e), (f), (g), and (h) without the corresponding labels. Note that some components of an object category are much similar to components in other object categories (for example, the new ball is much similar to the new cylinder captured from the cap side than to the old ball). Categorizing a unseen object based only on similarities of visual feature are useless in this case. However, if the system successfully learned the correspondence between object-oriented behaviors and labels, it should be able to associate the labels to the new objects.



(a) robot

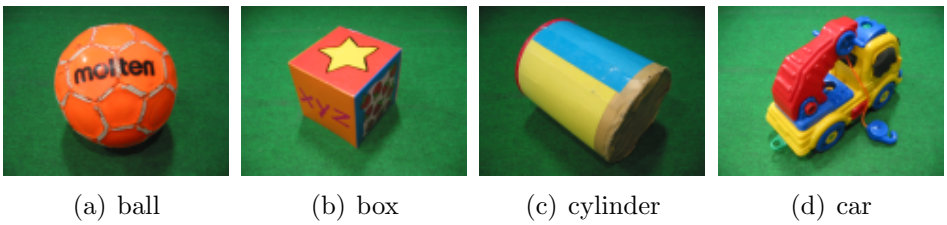


(b) CCD camera



(c) omni-directional wheels and kicking device

Figure 4.7: Mobile robot used in the experiment.



(a) ball (b) box (c) cylinder (d) car

(e) new ball (f) new box (g) new cylinder (h) new car

Figure 4.8: Objects used in the experiment.

4.3.2 Learner Setup

Hardware and Environment

The mobile robot utilized in the experiment is shown in Figure 4.7. The robot is equipped with a CCD camera (FireFly2 provided from Point Grey Reseach) with fixed view angle (Figure 4.7 (b)) and a kicking device in front with two arms that rotate independently in horizontal plane (Figures 4.7 (a), (c)). The robot is able to move into any direction in the horizontal plane by the use of omni-directional wheels (Figure 4.7 (c)). The robot interacted with the objects in a $3m \times 3m$ space covered with green color.

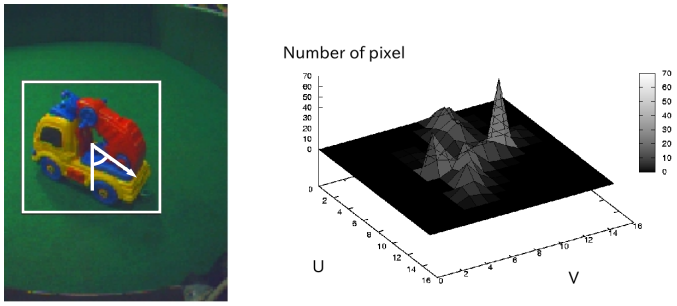
Visual information processing

A $240(H) \times 320(V)$ size image of objects was captured by a CCD camera and sent to laptop PC through IEEE1394 interface. Object region is extracted assuming that green color regions are backgrounds. Direction of principal axis is then calculated from the object region by principal component analysis as shown in Figure 4.9 (a) and utilized as state variable to run the object rolling behavior. On the other hand, we adopt YUV color space and UV color histogram as photometric feature. UV space is quantized into 16×16, so the color histogram is a 256 dimension vector representing frequency of quantized colors in UV space. Figure 4.9 (b) shows an example of the UV space histogram. The system uses χ^2-divergence as distance metric for categorization. The χ^2-divergence of two photometric feature vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are calculated as follows

$$d(\boldsymbol{a}, \boldsymbol{b}) = \chi^2(\boldsymbol{a}, \boldsymbol{b}) = \sum_i \frac{(a_i - b_i)^2}{a_i + b_i}. \tag{4.14}$$

Learning system

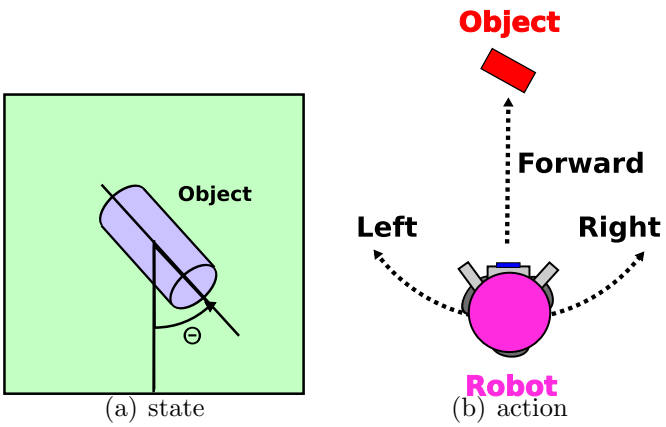
The state space for the learning modules of multi-module reinforcement learning system consists of direction of principal axis of the object $\theta \in [-90, 90]$ as shown in Figure 4.10(a). The state space is quantized into 7, and another state is added to represent a case when principal axis is uncertain. The learner is able to choose from three actions (Figure 4.10(b))



(a) extracted image (b) color histogram

Figure 4.9: Color histogram of object.

namely, kicking the object forward, moving clockwise and anticlockwise around the object. Finally, a reward whose value is proportional to moving distance of the object is given to the learner.



(a) state (b) action

Figure 4.10: State and action for task.

Parameters for Multi-module reinforcement learning system, adaptive networks, and Hebbian network are shown in Table 4.1.

Table 4.1: Parameters for the learning system.

Multi-module reinforcement learning system	
Decay factor γ	0.3
Threshold for action value error $DQ_{threshold}$	1.0
Adaptive network for behavior	
Learning rate β	1.0
Decay factor α	0.99
Initial value of network weight w_0	0.5
Adaptive network for label	
Learning rate β	1.0
Decay factor α	0.99
Initial value of network weight w_0	0.5
Hebbian network	
Increasing value of weight δ_{inc}	0.1
Decreasing value of weight δ_{dec}	0.1
Limit value of weight considered related W_{limit}	0.0

4.3.3 Experiment on object-oriented behavior learning and identification

The learner acquired the rolling behavior for the objects shown in Figures 4.8 (a), (b), (c) and (d). The assignments of new learning modules and changes of reliability of each module is shown in Figure 4.11, where curves with different line types correspond to different learning modules. The horizontal axis indicates the number of trials of interaction with the objects. One trial finished when the learner kicked the object 5 times, or when the learner lost the object. This indicates that if the object was to roll, approximately one successful kick is included in each trial. The learner experienced the objects in fixed order of ball, box, cylinder, and car. They interacted with the objects for 5 trials in the first 20 trials, for 3 trials in the next 12 trials, and after on, the object was switched after each trial. We introduced the scheduling to behavior learning process since if the learning modules were too immature, they would give large action value errors even though they were interacting with the corresponding objects. Taniguchi and Sawaragi [91] discuss this topic in detail. As shown in the figure, every time an unfamiliar object was introduced, the action value error exceeded the predefined limit $DQ_{threshold}(s, a) = 1.0$ (which makes the reliability less than

0) and a new learning module was assigned. The figure also shows that the same learning module is selected for each object. This shows that the system successfully acquired the set of learning modules that can be utilized to identify the objects based on object-oriented behaviors.

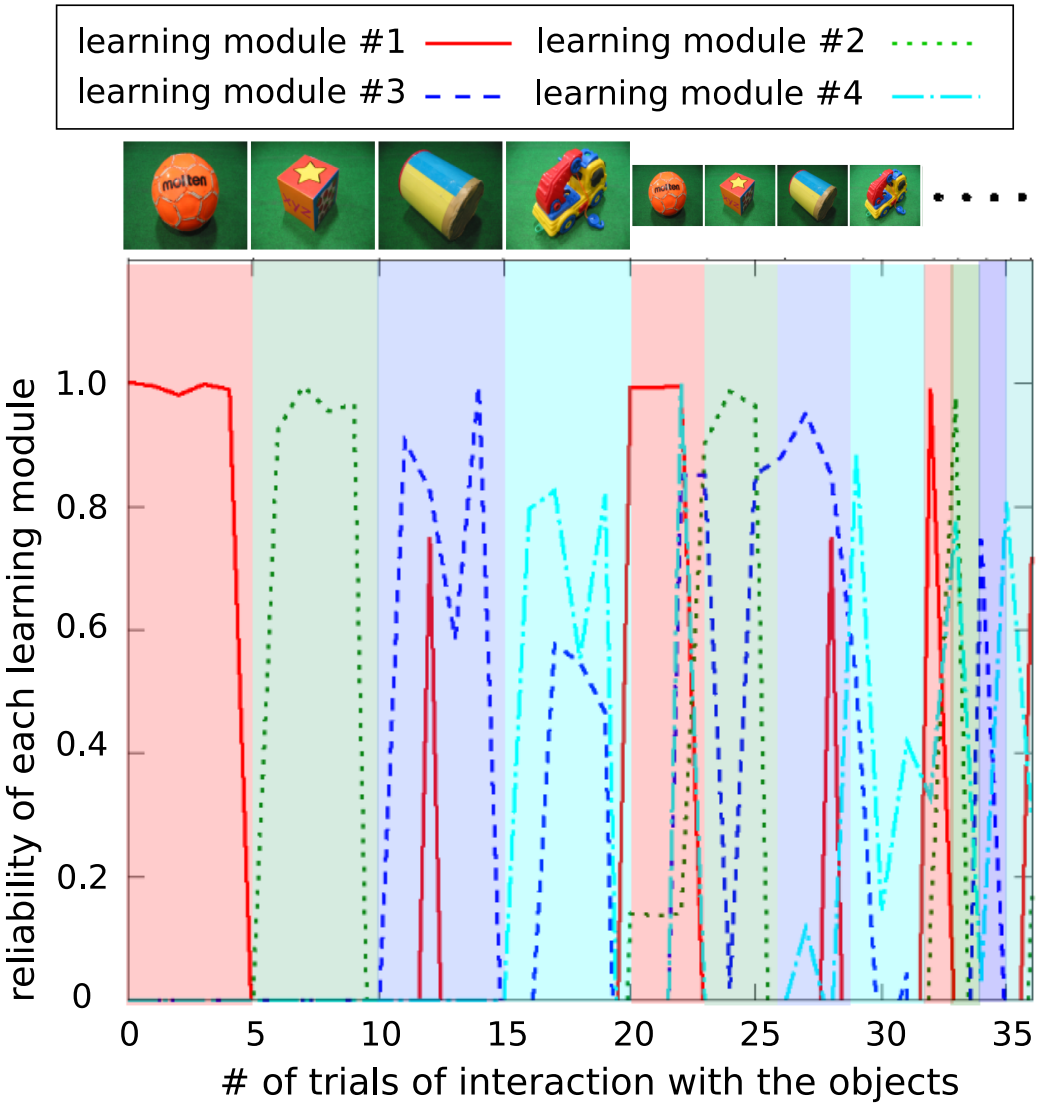
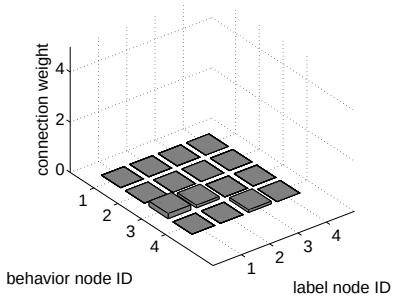


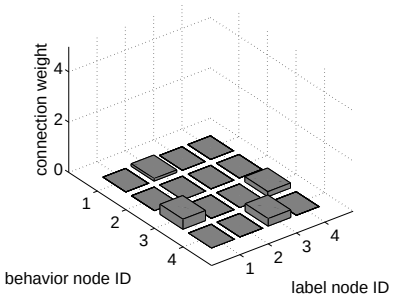
Figure 4.11: Action value error of each learning module while learning object-oriented behaviors.

4.3.4 Experiments on learning relation between object-oriented behaviors and labels

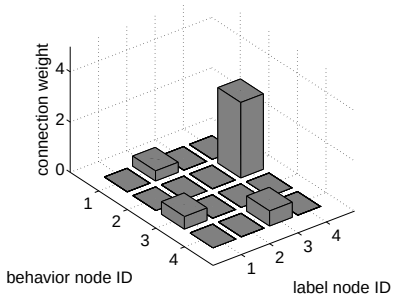
After the learner acquired the object-oriented behaviors and categorized the photometric feature space based on the behaviors, the caregiver taught the labels of the objects which the learner learned to roll (Figures 4.8 (a), (b), (c) and (d)) in random order. When the label was given, the system categorized the photometric feature space by assigning simultaneously given photometric features to the category of the given label. As the category of the labels develop, the relationship between object-oriented behaviors and labels is learned by modifying the weight of the Hebbian network. Figure 4.12 shows the Hebbian network's weight transition recorded from real robot experiment in accordance with the number of times the label was given. We can observe the system successfully learning the one-to-one correspondence of object-oriented behaviors and labels.



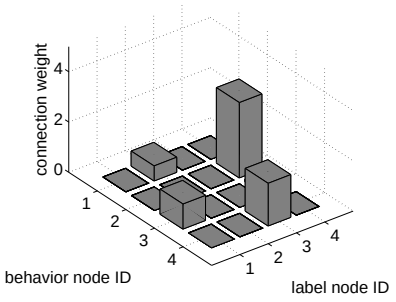
(a) 10 labels



(b) 20 labels



(c) 30 labels



(d) 60 labels

FIGURE 4.12: Weight transition of Hebbian network.

4.3.5 Experiment on label association

After learning the relation between object-oriented behaviors and labels, the learner was presented the new objects shown in Figures 4.8 (e), (f), (g) and (h) without the labels. Since the new objects afford object-oriented behaviors which have already been acquired previously, the learner was expected to be able to associate the learned labels to the new objects. Figure 4.13 shows the transition of success rate in labeling new objects as the learner interacted with the new objects and identified their object-oriented behaviors. The horizontal axis shows the number of trial of interaction the learner experienced with the new objects. During this interaction with the new objects, the multi-module reinforcement learning system continued the behavior learning process to adapt the learning modules to the new objects. The success rate of labeling shown in the vertical axis was calculated based on 400 sets of object image and label [2] produced in the environment of the experiment. Success rate of labeling is only about 20% at the start, which shows that existing methods are helpless for associating labels to new objects autonomously. As the learner interacted with the new objects, it expanded the categories of object-oriented behavior to photometric features of new objects. By this growth of the behavior based category, the learner became to answer appropriate labels to new objects following the process discussed in Section 4.2.6. After about 12 trials of interaction with the new objects, the mean success rate of labeling for all new objects reached 80 %. The success rate of labeling for the new car and the new box was not as high as the other two objects. This was due to the fact that the new car had a very similar color with the boxes. We can overcome this problem by introducing other photometric features such as edge histograms to form a better space for categorization.

[2] 100 sets of testing data was prepared for each object.

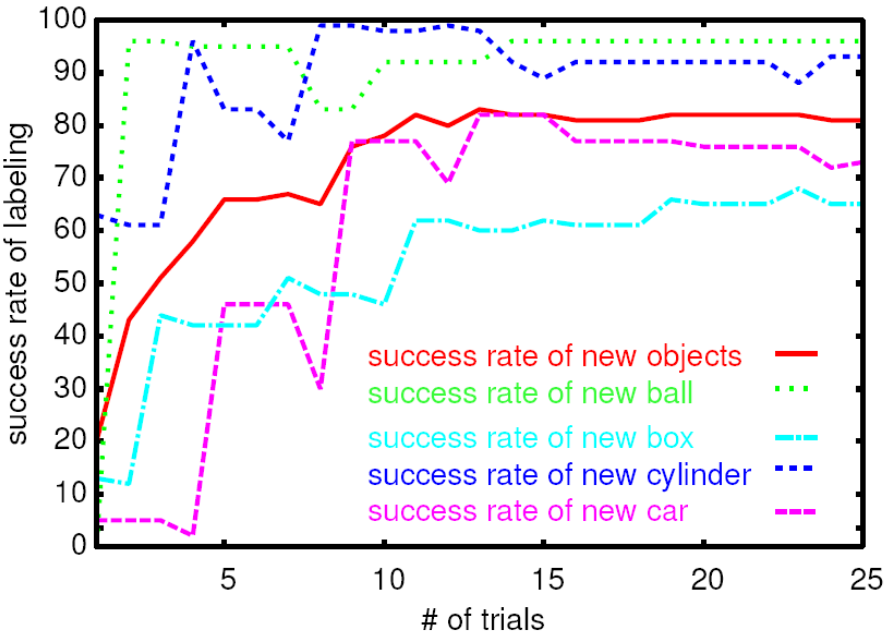


FIGURE 4.13: Association of labels to new objects.

4.4 Conclusion

We proposed a lexicon acquisition system which can associate labels to new objects with very different visual features according to the behavior that can be performed with the objects. The system was implemented to a robot learning labels about objects with different rolling preference. The robot learned the rolling behaviors for each object by reinforcement to successful kicks and identified the behaviors afforded by the objects through errors in state value prediction. Simultaneously, the robot also learned the correspondence of labels and behaviors through the overlap of visual features and successfully associated the labels to newly introduced objects. The proposed system can be considered as the model of the word generalization based on functions shown by Nelson et al. [17].

Chapter 5

Conclusion and future work

The dissertation investigated the role of physical agent-object interaction in the acquisition of daily object categories. Although computer vision studies have shown great increase in their performance, unsupervised learning of daily object categories still remains to be a difficult task. Inspired by the recent findings suggesting the role of physical interaction in object categorization, several experimental results were given to show how the physical exploration toward the objects could play a role in solving the task. In chapter 3, we introduced two experiments which reproduced infants' typical exploratory behaviors toward the objects; early manual touch and dynamic touch. The result of the first experiment showed that the manual behaviors, squeezing and tapping, could extract the surface stiffness of the objects which seems imporant for categorization through static and dynamic deformation of the manual skin. On the other hand, the second experiment showed that dynamic touch can ease the task of acquiring primitive object categories such as rigid objects, liquid, and paper material as deformable material by introducing auditory information processing of the cochlea. A system of lexicon acquisition based on object-oriented behavior learning was proposed and investigated in chapter 4. The system learned rolling behaviors as behavior modules based on reinforcement to the agent when it successfully kicked the objects. Then, by identifying the behaviors afforded by the objects through the errors of value prediction, the robot extended his categories and successfully generalized the labels to unfamiliar objects by uttering the corresponding labels to the behavior.

The long lasting issue addressed in chapter 2 should now be revisited. Is physical interaction required for object categorization? The works in chapter 3 showed that the design found in the human body helps extracting invariant properties of objects suitable for obtaining shared categories. In the case of the dynamic touch experiment, the auditory information processing found in the cochlea generated feature vectors suitable for primitive object categorization by ignoring information related to the size or contact conditions. We could come up with various other examples where the body helps in extracting information relative to object category acquisition. For example, heat property of objects can be measured by sensing the heat flow from the body to the object. Since the aim is to obtain humanlike categories, and those categories depend heavily on the structure of the human body, autonomous agents should also benefit with anthropomorphic design. One important feature of objects which are considered to be obtained only from physical interaction but essential for object categorization was the affordance of objects. The work based on behavior learning, in this respect, showed that the sensorimotor representation corresponding to object affordance can be obtained by a multi-module behavior learning architecture and used for object category acquisition.

Although we succeeded in giving examples on physical interaction playing a role in object categorization, the performance of the system was limited to a small set of object categories. The question as for future work is on how this could be increased. In case of the object categorization method based on behavior learning, there are some remaining issues such as the design of the value system and how the approach could be extended to obtain the complex object categories of humans. If the same algorithm is applied to a humanoid robot with the degree of freedom similar to that of humans, the robot will require too much time for learning the behaviors due to the increase of the search space. We could refer to the idea of embodiment for object categorization again. Humanlike body structures with compliant joints will enable the robots to effectively explore and obtain shared behaviors through the use of object dynamics. Once similar behavior sets are acquired, it would be easier to obtain shared categories through them. Extending the work to symbol grounding is another challenging issue to be addressed. Physical interaction with shared environment and body structures may then explain why humans share language structures within different cultures.

List of Publications

Articles in Journals

1. Shinya Takamuku, Yasutake Takahashi, and Minoru Asada. "Lexicon acquisition based on object-oriented behavior learning". Advanced Robotics. Vol.20. No.10. pp.1127-1145. 2006.

2. Shinya Takamuku and Ronald Arkin. "Multi-method learning and Assimilation". Robotics and Autonomous Systems. Vol. 55. Issue 8. 31. pp. 618-627. 2007.

3. Shinya Takamuku and Koh Hosoda and Minoru Asada. "Object Category Acquisition by Dynamic Touch". Advanced Robotics. (to appear)

Papers in Proceedings of International Conferences

1. Shinya Takamuku, Yasutake Takahashi, and Minoru Asada. "Lexicon Acquisition based on Behavior Learning". Proceedings of the 2005 4th IEEE International Conference on Development and Learning. 2005.

2. Shinya Takamuku, Gabriel Gomez, Koh Hosoda, Rolf Pfeifer. "Haptic discrimination of material properties by a robotic hand". Proceedings of the 2007 6th IEEE International Conference on Development and Learning. #76. 2007.

3. Shinya Takamuku, Koh Hosoda, and Minoru Asada. "Shaking eases Object Category Aquisition: Experiments with a Robotic Arm". Proceedings of the 7th International Conference on Epigenetic Robotics. 2007.

Papers in Symposiums and Meetings

1. Shinya Takamuku, Koh Hosoda, Minoru Asada. "Shaking eases Object Categorization". The 1st Asada Synergistic Intelligence Symposium. P01. 2007.

2. Shinya Takamuku, Yasutake Takahashi, Minoru Asada, "Lexicon Acquisition from Object-oriented Behavior Learning", in Poster Session of Interdisciplinary College. Guenne Germany. March. 2006.

References

[1] Y. Sakagami, R. Watanabe, and C. Aoyama. *The intelligent asimo: System overview and integration.* Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems pp. 2478–2483 (2002).

[2] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu. *Robovie: an interactive humanoid robot.* Industrial Robot **28**(6), 498 (2001).

[3] Y. Kuroki, M. Fujita, T. Ishida, K. Nagasaka, and J. Yamaguchi. *A small biped entertainment robot exploring attractive applications.* Proceedings of the 2003 IEEE International Conference on Robotics and Automation **1**, 471 (2003).

[4] K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi. *Humanoid robot hrp-2.* Proceedings of IEEE International Conference on Robotics and Automation pp. 1083–1090 (2004).

[5] M. Fujita and H. Kitano. *Development of a quadruped robot for robot entertainment.* Autonomous Robots **5**, 7 (1998).

[6] J. Forlizzi and C. DiSalvo. *Service robots in the domestic environment: a study of the roomba vacuum in the home.* Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction pp. 258–265 (2006).

[7] P. Pott, H. peter Scharf, and M. Schwarz. *Today's state of the art in surgical robotics.* Journal of Computer Aided Surgery **10**(2), 101 (2005).

[8] L. Steels. *The Talking Heads Experiment. Volume 1. Words and Meanings* (Laboratorium, Antwerpen, 1999).

[9] D. B. Lenat. *Cyc: A large-scale intestment in knowledge infrastructure.* Communications of the ACM **38**(11), 33 (1995).

[10] K. Furukawa, I. Kobayashi, T. Ozaki, and M. Imai. *A model of children's vocabulary acquisition using inductive logic programming.* In *Proceedings of the 2nd International Conference on Discovery Science*, pp. 321–322 (1999).

[11] S. Nolfi and D. Marocco. *Active perception: A sensiromotor account of object categorization.* Proceedings of the seventh international conference on simulation of adaptive behavior on From animals to animats pp. 266–271 (2002).

[12] G. Lakoff. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind* (Univ. of Chicago Press, 1987).

[13] J. Anglin. *Vocabulary development : A morphological analysis* (Monographs of the Society for Research in Child Development, 1993).

[14] W. Quine. *Word and Object* (MIT Press, 1960).

[15] E. M. Markman. *Categorization and Naming in Children: Problems of Induction* (MIT Press, Bradford Books, 1989).

[16] B. Landau, L. Smith, and S. Jones. *The importance of shape in early lexical learning.* Cognitive Development **3**, 299 (1988).

[17] K. Nelson, D. G., R. Russell, N. Duke, and K. Jones. *Two-year-olds will name artifacts by their functions.* Child Development **71**(5), 1271 (2000).

[18] H. Kobayashi. *The influence of adults' action on children's inferences about word meanings.* Japanese Psychological Research **41**(1), 35 (1999).

[19] G. Kreiman, C. Koch, and I. Fried. *Category-specific visual responses of single neurons in the human medial temporal lobe.* Nature Neuroscience **3**(9), 946 (2000).

[20] A. Slater and M. Lewis. *Introduction of Infant Development*, chap. 7 (Oxford Univ. Pr., 2007).

[21] J. Krichmar and G. Edelman. *Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device.* Cerebral Cortex **12**, 818 (2002).

[22] P. Rochat. *The Infant's World*, chap. 3 (Harvard Univ Press, 2004).

[23] J. J. Freyd. *The mental representation of movement when static stimuli are viewed.* Perception and psychophysics **33**, 575 (1983).

[24] C. Gerlach, I. Law, and O. Paulson. *When action turns into words. activation of motor-based knowledge during categorization of manipulable objects.* Journal of Cognitive Neuroscience **14**(8), 1230 (2002).

[25] J. Gibson. *The Ecological Approach to Visual Perception* (Lawrence Erlbaum Associates, London, 1986).

[26] E. J. Gibson. *Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge.* Annual Review of Psychology **39**, 1 (1988).

[27] E. Thelen and L. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT Press, A Bradford Book, 1994).

[28] G. Reeke and G. Edelman. *Selective networks and recognition automata.* Annals of the New York Academy of Sciences **426**, 181 (1984).

[29] R. Pfeifer and C. Scheier. *Understanding Intelligence*, chap. 12 (MIT Press, Bradford Books, 1999).

[30] M. Asada, K. MacDorman, H. Ishiguro, and Y. Kuniyoshi. *Cognitive developmental robotics as a new paradigm for the design of humanoid robots.* Robotics and Autonomous Systems **37**(2-3), 185 (2001).

[31] A. Pinz. *Object categorization.* Foundations and Trends in Computer Graphics and Vision **1**(4), 255 (2006).

[32] C. Harris and M. Stephens. *A combined conrner and edge detector.* Proc. 4th Alvey Vision Conference pp. 147–151 (1988).

[33] P. Beaudet. *Rotational invariant image operators.* Proc. International Conference on Pattern Recognition pp. 579–583 (1978).

[34] T. Kadir and M. Brady. *Scale, saliency and image description.* International Journal of Computer Vision **45**(2), 83 (2001).

[35] R. Laganiere. *Morphological corner detection.* Proc. of 6th International Conference on Computer Vision pp. 280–285 (1999).

[36] J. Matas, O. Chum, M. Urban, and T. Pajdla. *Robust wide baseline stereo from maximally stable extremal regions.* Proceedings of the 13th British Machine Vision Conference pp. 384–393 (2002).

[37] S. Smith and J. M. Brady. *Susan - a new approach to low level image processing.* International Journal on Computer Vision **23**(1), 45 (1997).

[38] K. Mikolajczyk and C. Schmid. *An affine invariant interesting point detector.* Proc. of Europian Conference on Computer Vision pp. 128–142 (2002).

[39] T. Kadir, A. Zisserman, and M. Brady. *An affine invariant salient region detector.* Proc. of Europian Conference on Computer Vision pp. 228–241 (2004).

[40] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. *Visual categorization with bags of keypoints.* Proc. of European Conference on Computer Vision pp. 1–22 (2004).

[41] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. *Generic object recognition with boosting.* IEEE Trans. on Pattern Analysis and Machine Intelligence **28**(3), 416 (2006).

[42] M. Burl, M. Weber, and P. Perona. *A probabilistic approach to object recognition using local photometry and global geometry.* Proc. of European Conference on Computer Vision pp. 628–641 (1998).

[43] R. Fergus, P. Perona, and A. Zisserman. *Object class recognition by unsupervised scale-invariant learning.* Proc. of International Conference on Computer Vision and Pattern Recognition pp. 264–271 (2003).

[44] R. Fergus, P. Perona, and A. Zisserman. *A visual category filter for google images.* Proc. of European Conference on Computer Vision pp. 242–256 (2004).

[45] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. *Learning object categories from google's image search.* Proc. of International Conference on Computer Vision (2000).

[46] J. Sivic, B. Russell, A. Efros, A. Zisserman, and W. Freeman. *Discovering objects and their location in images.* Proc. of International Conference on Computer Vision (2005).

[47] D. Roy and A. Pentland. *Learning words from sight and sounds: A computational model.* Cognitive Science **26**(1), 113 (2002).

[48] D. Moore, I. Essa, and M. Hayes. *Exploiting human actions and object context for recognition tasks.* Proc. in International Conference on Computer Vision pp. 80–86 (1999).

[49] M. Veloso, P. Rybski, and F. Hundelshausen. *Focus: A generalized method for object discovery for robots that observe and interact with humans.* Proc. in International Conference on Human Robot Interaction pp. 102–109 (2006).

[50] S. Lederman and R. Klatzky. *Haptic exploration and object representation.* Vision and Action: The Control of Grasping (1990).

[51] E. bushnell and P. Boudreau. *Motor development and the mind: The potential role of motor abilities as determinant of aspects of perceptual development.* Child Development **64**(4), 1005 (1993).

[52] Y. Tada, K. Hosoda, and M. Asada. *Sensing ability of anthropomorphic fingertip with multi-modal sensors.* Proc. of Internaional Conference on Intelligent Autonomous Systems (2004).

[53] L. Natale, G. Metta, and G. Sandini. *Learning haptic representation of objects.* Proc. of International Conference on Intelligent Manipulation and Grasping (2004).

[54] L. Natale and E. Torres-Jara. *A sensitive approach to grasping.* Proc. of International Workshop on Epigenetic Robotics (2006).

[55] E. Torres-Jara, L. Natale, and P. Fitzpatrick. *Tapping into touch.* Proc. of International Workshop on Epigenetic Robotics (2005).

[56] T. Ogata, H. Ohba, K. Komatani, J. Tani, and H. G. Okuno. *Extracting multi-modal dynamics of objects using rnnpb.* Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems pp. 160–165 (2005).

[57] C. G. Atkeson, C. H. An, and J. M. Hollerbach. *Rigid body load identification for manipulators.* Proceedings of 24th Conference on Decision and Control pp. 996–1002 (1985).

[58] M. Suzuki, K. Noda, Y. Suga, T. Ogata, and S. Sugano. *Dynamic perception after visually guided grasping by a human-like autonomous robot.* Advanced Robotics **20**(2), 233 (2006).

[59] T. Nagai and N. Iwahashi. *Object categorization using multimodal information.* Proc. of IEEE region 10 conference of TENCON pp. 1–4 (2006).

[60] S. Nolfi. *Adaptation as a more powerful tool than decomposition and integration.* Proc. of International Conference on Machine Learning (1996).

[61] S. Nolfi and D. Marocco. *Active perception: A sensorimotor account of object categorization.* From animals to animats: Proc. of International Conference on Simulation of Adaptive Behavior (2002).

[62] R. Beer. *Toward the evolution of dynamical neural networks for minimally cognitive behavior.* From animals to animats: Proc. of International Conference on Simulation of Adaptive Behavior (1996).

[63] C. Scheier and R. Pfeifer. *Classification as sensory-motor coordination: A case study on autonomous agents.* In Proc. of European Conference on Advances in Artificial Life (1995).

[64] A. Streri and J. Feron. *The development of haptic abilities in very young infants: From perception to cognition.* Infant Behavior and Development **28**(3), 290 (2005).

[65] F. Jouen and M. Molina. *Exploration of the newborn's manual activity: A window onto early cognitive processes.* Infant Behavior and Development **28**(3), 227 (2005).

[66] H. Yokoi, A. Arieta, R. Katoh, W. Yu, I. Watanabe, and M. Maruishi. *Mutual adaptation in a prosthetics application in embodied artificial intelligence.* Lecture Notes in Computer Science **3139** (2004).

[67] Y. Ishikawa, W. Yu, H. Yokoi, and Y. Kakazu. *Research on the double power mechanism of the tendon driven robot hand.* The Robotics Society of Japan pp. 933–934 (1999).

[68] K. Hosoda, Y. Tada, and M. Asada. *Anthropomorphic robotic soft fingertip with randomly distributed receptors.* Robotics and AUtonomous Systems **54**(2), 104 (2006).

[69] J. N. Hoffmann, A. G. Montag, and N. J. Dominy. *Meissner corpuscles and somatosensory acuity: The prehensile appendages of primates and elephants.* The anatomical record part A **281A**, 1138 (2004).

[70] T. Kohonen, J. Hynninen, J. Kangas, and J. Laaksonen. *Som pak: The self-organizing map program package.* Helsinki University of Technology Technical Report **A31** (1996).

[71] M. Turvey. *Dynamic touch.* American Psychologist **51**, 1134 (1996).

[72] M. M. Williamson. *Rhythmic robot arm control using oscillators.* Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems pp. 77–83 (1998).

[73] M. T. Turvey. *Dynamic touch.* American Psychologist **51**, 1134 (1996).

[74] T. Shimizu and H. Norimatsu. *Detection of invariants by haptic touch across age groups: rod-length perception.* Perceptual and Motor Skills **100**, 543 (2005).

[75] H. Kloos and E. Amazeen. *Perceiving heaviness by dynamic touch: An investigation of the size-weight illusion in preschoolers.* British journal of developmental psychology **20**(2), 171 (2002).

[76] M. Suzuki, K. Noda, Y. Suga, T. Ogata, and S. Sugano. *Dynamic perception after visually guided grasping by a human-like autonomous robot.* Advanced Robotics **20**(2), 233 (2006).

[77] M. M. Williamson. *Rhythmic robot arm control using oscillators.* In Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems pp. 77–83 (1998).

[78] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella. *Towards a model for tool-body assimilation and adaptive tool-use.* Proceedings of the International Conference on Development and Learning (2007).

[79] A. Fregolent and A. Sestieri. *Identification of rigid body inertia properties from experimental data.* Mechanical Systems and Signal Processing **10**(6), 697 (1996).

[80] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. Lamantia, J. O. McNamara, and S. M. Williams. *Neuroscience*, chap. 13 (Sinauer Associates Inc., 2004).

[81] T. Kohonen. *Self-Organizing Maps* (Springer-Verlag, 1995).

[82] G. Lakoff and M. Johnson. *Metaphors we live by* (Univ. of Chicago Press, 1980).

[83] S. Harnad. *The symbol grounding problem.* Physica D **42**, 335 (1990).

[84] Y. Sugita and J. Tani. *Learning semantic combinatoriality from the interaction between linguistic and behavioral processes.* Adaptive Behavior **13**(1), 33 (2005).

[85] D. Wolpert and M. Kawato. *Multiple paired forward and inverse models for motor control.* Neural Networks **11**, 1317 (1998).

[86] D. Milner and M. Goodale. *The visual brain in action* (Oxford University Press, 1995).

[87] P. Fitzpatrick and G. Metta. *Early integration of vision and manipulation.* Adaptive Behavior **11**(2), 109 (2003).

[88] C. Watkins and P. Dayan. *Technical note: Q-learning.* Machine Learning **8**, 279 (1992).

[89] M. Hassoun. *Fundamental of Artificial Neural Networks* (MIT Press, 1995).

[90] L. Steels and T. Belpaeme. *Coordinating perceptually grounded categories through language: A case study for colour.* Behavioral and Brain Sciences **28**, 469 (2005).

[91] T. Taniguchi and T. Sawaragi. *Assimilation and accomodation for self-organzational learning of autonomous robots' a proposal of dual-schemata model*. Prof. of International Symposium on Computational Intelligence in Robotics and Automation pp. 277–282 (2003).

Appendix

Table 5.1: Labels and coresponding conditions for the shaking experiment.

label	condition
rigid(A)	Aluminum stick with 25cm length
rigid(B)	Aluminum stick with 50cm length
rigid(C)	Ferrum stick with 25cm length
rigid(D)	A wrench
rigid(E)	A file
rigid(F)	PET bottle without water
rigid(G)	Aluminum stick with 25cm length held in perpendicular direction
rigid(H)	Aluminum stick with 25cm length shaken with half control frequency (0.5Hz)
paper(A)	5 pages of A4 paper
paper(B)	5 pages of B4 paper
paper(C)	1 page of A4 paper
paper(D)	A magazine
paper(E)	A newspaper
paper(F)	A A6 notebook with 100 pages
paper(G)	5 pages of A4 paper held in perpendicular direction
paper(H)	5 pages of A4 paper shaken with half control frequency (0.5Hz)
water(A)	500ml PET bottle with 100g water
water(B)	500ml PET bottle with 200g water
water(C)	A tall 1.0l PET bottle with 100g water
water(D)	500ml PET bottle with 100g water and hexagonal cross section shape
water(E)	350ml Aluminum can with 100g water
water(F)	A short 1.0l PET bottle with 100g water
water(G)	500ml PET bottle with 100g water held upside down
water(H)	500ml PET bottle with 100g water shaken with half control frequency (0.5Hz)