Developmental Model of Joint Attention by Utilizing Contingency in Interaction

（相互作用の随伴性を利用した共同注意発達モデル）

A dissertation submitted to the Department of Adaptive Machine Systems
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Engineering
at the
OSAKA UNIVERSITY

Hidenobu Sumioka

July 2008

Thesis Supervisor: Minoru Asada
Title: Department of Adaptive Machine Systems,
Graduate School of Engineering, Osaka University

Thesis Committee: Minoru Asada, Chair
Hiroshi Ishiguro
Koh Hosoda

Abstract

Joint attention, defined as looking at the same object that someone else is looking at, plays an important role for imitation learning, language communication, and mind-reading in human and human-robot communication. In this study, we aim at building a robot that acquires various forms of joint attentional behavior in an infant-like manner in order to provide a new understanding of the developmental process of joint attention, and to realize the function of joint attention in a robot. We focus on contingency in interaction between a caregiver and a robot to make the robot acquire various forms of behavior. This causes two problems: how a robot can find a contingent relationship in interaction with its environment including a caregiver, and what kind of learning mechanism and environmental setting a robot requires to sequentially acquire several forms of joint attentional behavior based on the found contingencies. In order to address these problems, we deal with the following three issues.

First, an information theoretic measure is proposed to find a contingency structure in the interaction between a caregiver and a robot. We investigate how the proposed measure is used to quantify the contingency inherent in computer simulations of face-to-face interaction. The results indicate that it enables a robot to find a contingent relationship between face pattern of a caregiver and shifting its own gaze for learning a sensory-motor map to achieve gaze following. Next, we propose a learning mechanism that iteratively acquires several kinds of joint attentional behavior based on the proposed measure. The mechanism constructs a sensory-motor mapping to perform the behavior that reproduces the found contingency. It uses outputs from the previously acquired sensory-motor maps to change the contingency structure of the interaction with a caregiver. The results of computer simulations indicate that a robot acquires a series of actions related to joint attention in an order that almost

matches with an infant's development of joint attention. Furthermore, in order to apply the proposed mechanism to a real robot, the real-time face-to-face interaction where a human caregiver and a robot shift their gaze without assuming turn-taking of gaze change between them is considered. How a robot decides when to shift its gaze to acquire gaze following with a human is addressed. We propose a method that solves the issue by introducing an attention selector based on a measure consisting of saliencies of object features and motion information. The motion cues are expected to reduce the number of incorrect training data pairs due to asynchronous interaction that affects the convergence of the contingency learning. The experimental result shows that gaze shift utilizing motion cues enables a robot to synchronize its own motion with human motion and to learn gaze following efficiently. Finally, this dissertation is concluded and future issues are given.

Acknowledgments

First of all, I express my gratitude to Professor Minoru Asada for his support, guidance, and encouragement. He guided me to the exciting and challenging field of cognitive developmental robotics. His critical and valuable comments enabled me to realize not only the lack in this study but also what I should learn. I would not be able to accomplish this dissertation without them.

I am also very grateful to associate professor Koh Hosoda. His comments from an engineering viewpoint have always given me an opportunity to review this study from the viewpoint. I learned important things not only to advance my own research but also to spend happy life as a researcher. His advice has helped me shape the way I think.

Professor Hiroshi Ishiguro has spared his valuable time to discuss about this study. He has also given meaningful advice about this study. I'm very grateful to him.

I also thank all the present and the past members in Emergent Robotics and Adaptive Robotics Laboratories at Osaka University.

I express my gratitude to Dr. Yuichiro Yoshikawa. He has always supported me since I joined Emergent Robotics Laboratory. He has given me meaningful advice to solve many problems about my study and life. I was so lucky to have an opportunity to discuss with him. Through the discussion, I was able to learn various things.

I thank Dr. Yasutake Takahashi and Dr. Yukie Nagai. Dr. Yasutake Takahashi has given me a chance to learn basic image processing through his program. I would not perform the experiments where I used a real robot without the program. Dr. Yukie Nagai is a great senior researcher of my research topic and has provided me with knowledge about the developmental process of infants. I was influenced by her active

attitude of opening a new research field in foreign country. I would not try to experience studying abroad in Germany without seeing her attitude. I thank Dr. Takashi Takuma and Dr. Yasunori Tada. Dr. Takashi Takuma has instructed how to design a robot. Dr. Yasunori Tada has supported me for not only my study but also my works as system and database administrator in our laboratories. I also thank Dr. Masaki Ogino, Mr. Shinya Takamuku, and Mr. Katsushi Miura. They have given me not only valuable comments on my study but also a good time in the laboratories. I also thank Mr. Joschka Boedecker. He spared his valuable time to proofread this dissertation and gave me meaningful advice. Other members who are not described here also supported me in various ways. Thank you very much.

Finally, I would like to say “ありがとう” to my family. They have supported me in every way. My parents, Nobuo and Mayumi, have allowed me to spend a long student life. They have kept supporting and encouraging me though they had little knowledge about my research. My brother and sister, Kazuhiro and Rieko, have supported me in my daily life. They gave me some space to relax. Thank you very much for your love.

July 7, 2008
Hidenobu Sumioka

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Communicative robots have received increasing attention in autonomous robotics since communication is one of the most fundamental functions in both humans and robots [1; 2; 3]. They are expected to perform complicated collaborative tasks with human partners by knowing what to do, and how to do it, through communication with their partners. In such situations, sharing attention with their partners is the first critical step for the robots to realize smooth communication.

Humans, however, do not confront such a difficult problem. When we learn something from another person, we can find what to learn by sharing the person's attention from the observation of gaze direction and gesture. We can even estimate the person's intention by attention sharing. Therefore, attention sharing is a basic process to understand the other's intention [4; 5]. Human infants seem to acquire this ability for attention sharing through interaction with their caregivers since they do not have such a ability from the beginning. They can perform imitation learning, understand the other's intention, and acquire language after they acquire understanding, manipulating, and coordinating other's attention [5]. Therefore, understanding when and how they acquire this ability is one of the central topics in developmental psychology, cognitive science, and neuroscience.

Human gaze direction provides useful information to understand the direction of the other's attention. Joint attention, especially joint visual attention defined as

looking at an object that someone else is looking at, is regarded as the elemental component of attention sharing [6]. Therefore, developmental psychologists have investigated the development of various forms of behavior to achieve joint attention [7]. There are two types of abilities to achieve joint attention:

- following gaze of others and/or pointing gesture and,
- directing gaze of others to something to be looked at,

where the former is called RJA (responding to joint attention) and the latter is named as IJA (initiating joint attention) [8; 9]. The abilities of joint attention in human infants appears to develop from RJA to IJA. First, infants begin to follow gaze of others from about six months of age [10]. After that, they show gaze alternation (successive looking between a caregiver and an object), social referencing (looking back at a caregiver to obtain her emotional information in an unknown situation), showing an object, and pointing to an object in an almost fixed order between about nine and twelve months of age [11].

Some psychologists have claimed that human beings have an innate ability of attention sharing independent from the development of joint attentional behavior [12] while others have hypothesized that the development of these forms of joint attentional behavior is related to that of the ability of attention sharing [5; 13]. In one of these hypotheses, Moore and his colleagues mentioned that infants can acquire joint attentional behavior by learning based on a contingent structure of interaction with a caregiver, such as giving a reward to infants only when they achieve gaze following for a caregiver [13; 14]. In addition, they also suggested that the experience of engaging in joint attention leads infants to understand sharing attention with others once they begin to show the acquired behavior. However, Tomasello pointed out that they do not indicate how learning based on the contingent structure produces the developmental order of joint attention [5]. He also described that longitudinal studies on several forms of behavior to achieve joint attention should be conducted to verify those hypotheses. In practise, a longitudinal study has started in Japan [15], but it is unclear why such developmental order is produced.

In AI (artificial intelligence) and robotics, synthetic approaches have been applied

to understand development of joint attention from the viewpoint of cognitive developmental robotics [16] aiming at understanding the cognitive developmental process of biological agents and realizing cognitive functions in robots. Synthetic approaches [17; 16] are expected to provide new knowledge about problems that cannot be dealt with through analytical methodology of existing disciplines, such as developmental psychology and cognitive science. It aims at acquiring new knowledge about internal mechanisms of a biological system through the repetition of modeling the system based on knowledge in the existing disciplines, verifying the model, and improving it. Synthetic approaches are expected to help the advancement of existing disciplines as well as building more intelligent system. Some synthetic studies focusing on development of gaze following have indicated that contingency in interaction with a caregiver enables a robot to acquire gaze following [18; 19; 20]. However, it still remains unclear what kind of underlying mechanism determines the development order of several forms of the joint attentional behavior.

Therefore, this dissertation addresses the issue of building a mechanism that enables a robot to autonomously acquire several forms of behavior to achieve joint attention through interaction with a caregiver in an infant-like manner. In accordance with Moore's hypothesis, we focus on contingency in interaction with a caregiver as a clue for acquisition of joint attentional behavior. A robot finds the contingency by itself and performs actions to reproduce the found contingency. Through the repetition of finding the contingency and its reproduction, the robot acquires different forms of behavior. We expect that the proposed mechanism does not only provide new knowledge about the developmental relationships between different forms of joint attentional behavior, but also enables a robot to develop socially in an infant-like manner.

1.2 The Problems

We consider building a learning mechanism that enables a robot to autonomously acquire several forms of joint attentional behavior based on contingency of the interaction with a caregiver in an infant-like manner. Here, contingency means relationship among a preceding stimulus, an action, and its consequence. If a response of a robot

to a stimulus causes a certain result with high probability, the contingency exists between the response and stimulus on the result.

In a naturalistic interaction between a caregiver and an infant, the caregiver does not necessarily provide any explicit instruction for the infant to ease acquisition of joint attention related actions. The infant should find contingency through interaction with the caregiver. In order to realize a robot that acquires the actions autonomously, therefore, we need to address two issues: how does the robot find the contingency in the interaction and how can it autonomously acquire various forms of behavior based on that.

Finding contingency In previous synthetic studies on development of gaze following [18; 19; 20; 21], the designers provided a robot with a priori knowledge about what kind of contingency should be used to acquire gaze following. All that the robot does is to learn the predefined sensory-motor map. When the robot does not have such knowledge, however, it first needs to find a sensory-motor map including an appropriate contingency by itself. To address this issue, we introduce a contingency measure to find contingency in interaction. The robot decides sensory-motor map that it should learn based on the measure.

Moreover, we should consider that contingency in interaction with a caregiver is influenced by when a robot acts. The robot cannot find the contingency unless it coordinates its behavior adequately. This is a problem which we call *asynchronous problem* for finding contingent structure. The robot often encounters this problem in real-time interaction. For example, in face-to-face interaction between a human and a robot, the robot can find the contingency involving the human gaze easily under the assumption of turn taking of their gaze changes because it can knows when the human is looking at something. In the case of autonomous gaze shifting, however, that is difficult because the robot has to detect when the human is looking at the object.

We hypothesize that innate bias of human for shifting the attention helps human infants to solve this problem. Therefore, we introduce a human-like attention system. Motion information in the environment is utilized because it is expected to provide information about when a human begins to shift the gaze and where the human

directs the attention.

Autonomously acquiring various actions based on finding and reproducing contingency In order to make a robot acquire several forms of joint attentional behavior without any explicit instruction from the caregiver, we focus on reproducing the found contingency. We do this because human infants seem to not only find contingency in interaction with their caregivers but also try to perform the behavior to reproduce the found contingency [22]. We hypothesize that reproducing the found contingency further leads novel contingency to emerge from interaction with the caregiver because it introduces a change of the caregiver's response to the robot into the interaction (Figure 1.1). We expect this loop of finding and reproducing the contingency to enable autonomous development of joint attention.

Therefore, we aim at building a robot that can acquire various forms of joint attentional behavior through finding contingency in interaction with a caregiver and reproducing the found contingency.

1.3 Issues to be tackled

To build a learning mechanism that enables a robot to autonomously acquire several forms of joint attentional behavior, we deal with the components mentioned above. The proposed mechanism is examined in a computer-simulated setting involving face-to-face interaction. In addition, to apply the mechanism to a real robot interacting with a human caregiver naturally, real-time interaction without any synchronization assumption is considered. We deal with the problem of how to decide when to shift the gaze to achieve gaze following with a human, that is the asynchronous problem. Figure 1.2 represents what types of problems are addressed in each chapter. The rest of the dissertation is organized as follows:

Chapter 2 The findings about the development of various forms of joint attentional behavior from cognitive developmental science are described. We review them from three points of view: developmental process, its underlying mechanism, and the difference between individuals. In addition, previous synthetic studies on development

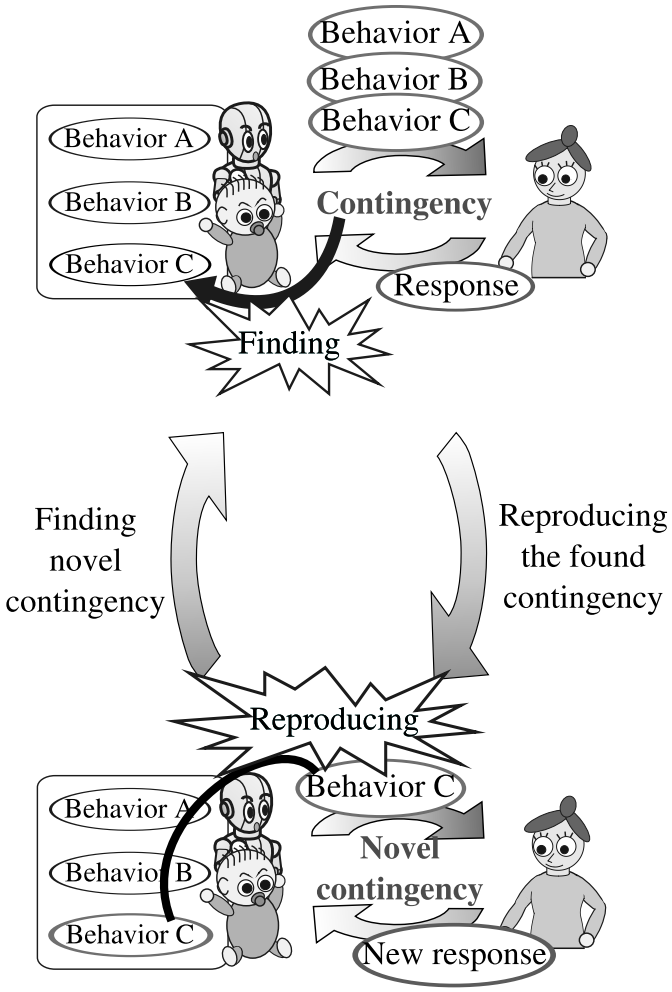


Figure 1.1: Autonomous acquisition of joint attentional behavior through finding contingency and its reproduction: first, a robot finds the contingency in interaction with a caregiver through performing several forms of behavior and observing the caregiver's response to them. Then, once it performs the behavior to reproduce the found contingency, the caregiver is expected to show new response. As a result, the robot has a chance to find novel contingency.

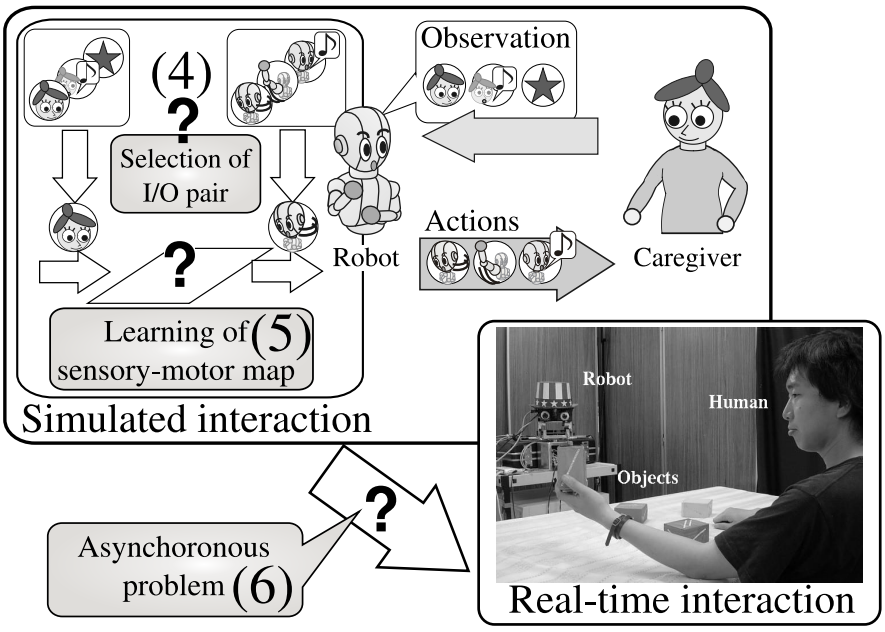


Figure 1.2: The model of autonomously social development. The gray boxes are addressed in this dissertation. The numbers correspond to the chapters where the problems related to learning based on contingency are addressed. Finding contingent pair between sensory information and action is tackled in chapter 4. Autonomous acquisition of contingent action based on the found pair is a main topic in chapter 5. Asynchronous problem is addressed in chapter 6 to use a learning mechanism proposed in chapter 5 in real-time interaction.

of joint attention and learning mechanisms for open-ended development are described.

Chapter 3 The face-to-face interaction between a robot and a human caregiver for development of joint attention is given. The robot observes environmental information and its own actions as variables. We explain about contingency in face-to-face interaction from the standpoint of a statistical bias in the combination of the variables.

Chapter 4 To find the contingent structure in face-to-face interaction between a caregiver and a robot, we propose an information theoretic measure that detects causality. *Transfer entropy* that shares some of the desired properties of mutual information but also takes into account the dynamics of information transport [23] is extended and utilized for that purpose.

We investigate how the proposed measure is used to quantify the contingency inherent in face-to-face interaction. In computer simulations of human-robot interaction, we examine which pair of perceptions and actions is selected as the contingent pair and show that the selected pairs can be used for learning a sensory-motor map for gaze following.

Chapter 5 To build a learning mechanism that enables a robot to autonomously acquire several forms of joint attentional behavior based on contingency in the interaction with a caregiver, we propose a mechanism that iteratively acquires several kinds of behavior based on the measure proposed in Chapter 4. The mechanism not only finds a combination of contingent variables but also constructs a sensory-motor mapping to reproduce behavior based on the found contingency. In the iterative process, a new variable expressing whether each sensory-motor mapping is used is added to promote finding the contingency depending on other contingent structures.

In computer simulations, we examine what kinds of actions related to joint attention can be acquired in order by changing the actions of the caregiver agent. The results indicate that a robot acquires a series of actions related to joint attention in an order that almost matches with an infant's development of joint attention. The difference between them is discussed based on the analysis of the robot behavior and

future issues are given.

Chapter 6 In order to adopt the proposed mechanism to a real robot, we address real-time interaction including the asynchronous problem. The issue is how to decide when to shift the gaze to achieve gaze following with a human.

We propose a method that solves the issue by introducing an attention selector based on a measure consisting of saliencies of object features and motion information. In order to realize natural interaction, that means real-time response without constrained synchronization of gaze shift between human and robot, self-organizing map (SOM) for real-time face pattern discrimination [24] and contingency learning for gaze following without external evaluation are utilized. The attention selector controls the robot gaze to switch often from the human face to an object and vice versa, and pairs of a face pattern and a gaze motor command are input to the contingency learning. The motion cues are expected to reduce the number of incorrect training data pairs due to asynchronous interaction that affects the convergence of contingency learning [19].

The experimental result shows that gaze shift utilizing motion cues enables a robot to synchronize its own motion with human motion and to learn gaze following efficiently in about 20 minutes.

Chapter 7 Finally, conclusions of this dissertation and future works are given to apply the proposed mechanism to a real robot.

Chapter 2

Related Work

Many researchers in different fields have investigated the developmental process of joint attention from different viewpoints. In this chapter, we review studies on the development of joint attention in not only developmental psychology and neuroscience but also robotics. The aims of this review are to make clear in what order the proposed mechanism is expected to acquire various forms of joint attentional behavior, and the difference between previous synthetic studies and our study.

2.1 Joint Attention

Joint attention is simply defined as looking where someone else is looking [10]. Many researchers in developmental psychology have mainly investigated the development of following the gaze of others [10; 14; 25] since Scaife and Bruner [26] reported that human infants show that capability before their first birthday. It is suggested that joint attention is one of the building blocks for social capabilities such as language communication [27; 28; 29; 30] and mind-reading [11; 12]. Therefore, studies on joint attention have recently been advanced taking into account other behaviors to achieve joint attention, such as following pointing of others (point following), successive looking between a caregiver and an object (gaze alternation), and pointing, as well as gaze following [7].

In this section, we review human development of joint attention from the following three viewpoints:

- when infants acquire them (development of joint attention);
- how infants acquire them (underlying mechanism of development of joint attention); and
- what differences there are among infants (developmental differences between individuals).

Section 2.1.1 shows in what order a robot should acquire various forms of joint attentional behavior. We indicate two main streams about the underlying mechanism of the development in section 2.1.2. Finally, we describe autism as one of the developmental disorders in section 2.1.3.

2.1.1 Development of joint attention

Figure 2.1 illustrates the typical development of joint attention related actions. In this section, we give an overview of the development of joint attentional actions during two developmental periods; the first 9 months of life when poor skills to achieve joint attention begin to emerge, and the period from 9 to 18 months when infants begin to follow and direct the attention and behavior of other persons.

The first nine months of life

From birth, human infants have some innate capacities such as preference to the human face [31], and objects with complex textures or symmetrical patterns [32]. They are also sensitive to eye-contact [33] and causality in their environment such as physical rules [34].

Around 6 month, they begin to follow the gaze of others. This behavior, however, seems responsive: they follow the gaze direction of their caregivers only when their caregivers are looking at an object in their visual field. In addition, they look at first distractor objects along their scan path. This earliest gaze following is called "ecological" mechanism [10]. When caregivers show a pointing gesture, six-month-old infants often look at the caregivers' finger or hand [35]. Infants in this period seem to shift their gaze reflexively by a cue of moving hands of their caregivers [36] and emotional change [37]. They do not appear to understand triadic relations among

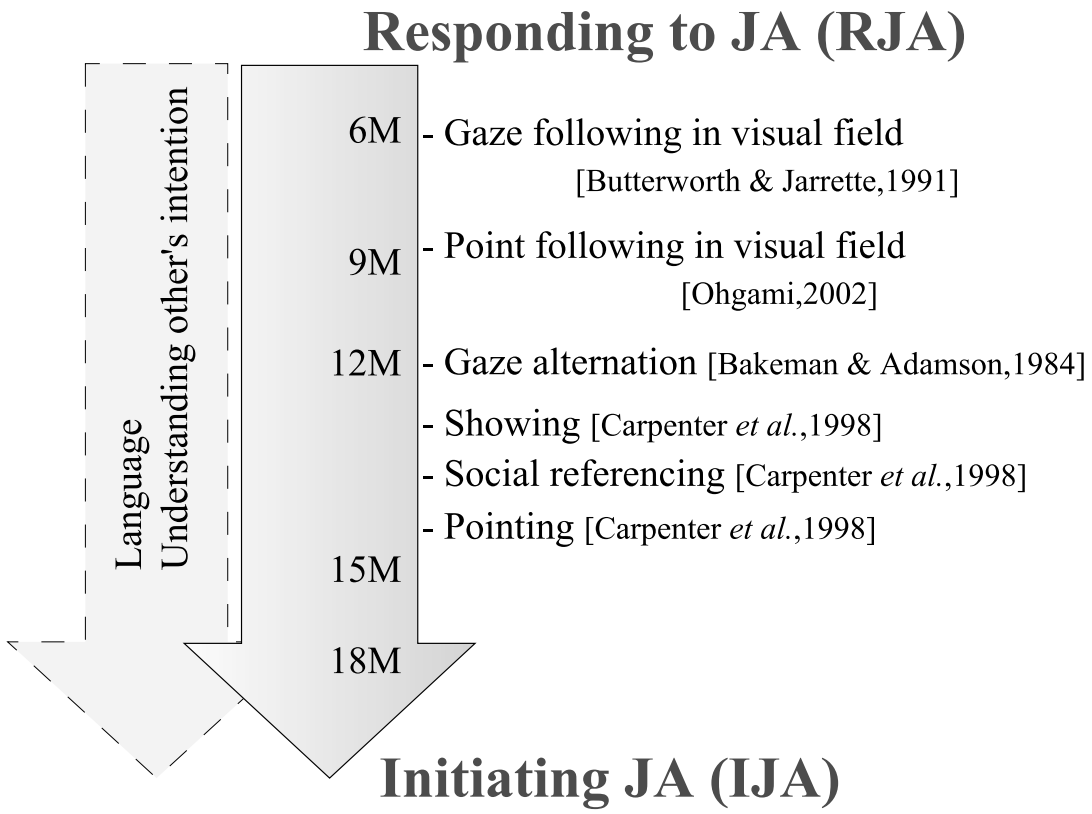


Figure 2.1: Development of joint attention related actions

their caregivers, an object, and themselves because they do not frequently look back at their caregivers after looking at an object.

After "the nine-month revolution"

Infants experience critical development occurring at around nine months old. This is called "the nine-month revolution" [5; 38]. During this period, they improve already acquired skills of joint attention dramatically and acquire different forms of joint attentional behavior. Twelve-month-old infants attain correctly following the gaze of others or pointing gesture. Butterworth [10] called this form of gaze following "geometric" mechanism. They appear to notice the attention of others [39]. There are observations that they can utilize the posture, pointing gesture, and motion information of their caregivers to achieve gaze following [25; 40]. At around 18-months old, they can follow the other's gaze to places outside their visual field. This is called "representational" mechanism [10].

Infants acquire several forms of behavior categorized as IJA such as alternately gazing between their caregivers and an object, and pointing. Infants at around twelve months perform showing, social referencing, i.e., looking back at their caregivers to obtain information about their emotion, and pointing [11]. They become interested in words around the same time and pay attention to a novel object when adults are talking about it [28]. Infants come to be engaged in triadic interaction. Bakeman and Adamson found 13-month-old infants kept gazing at their caregivers and at an object alternately [41].

As described above, the actions related to joint attention are acquired sequentially. Tomasello [5] describes that, in order to understand how infants develop these actions, it is important to investigate the underlying mechanism. Some studies about the development of joint attention have been conducted [15; 42].

The main purpose of this study is building a learning mechanism that enables a robot to autonomously acquire several forms of joint attentional behavior to understand the relations in their developmental processes as shown in Figure 1.2. In chapter 5, we discuss the difference between the results of computer simulations and an infant's development of joint attention.

2.1.2 The underlying mechanism of development of joint attention

There are mainly two theories about the underlying mechanism of developmental processes of joint attentional skills. Tomasello has mentioned that the mechanism is based on understanding others as intentional agents [5; 11] as suggested by observations that nine-months-old infants understand the purpose of CG (computer graphics) agents [43] and thirteen-months-old ones perform imitation learning [42].

On the other hand, Moore and Corkum have proposed that infants acquire the skills by learning based on contingent structure in interaction with a caregiver [13]. They also suggested that understanding of the other's intention progresses through the experience of engaging in joint attention [14]. There are some observations that support this hypothesis. Infants can not only find social contingency in interaction with a partner but also coordinate their own actions based on the found contingency [44; 45; 46; 47]. In addition, their caregivers show a contingent response to infant's behavior [22; 41; 48], and change according to the development of joint attention [49; 50; 51].

We consider that social contingency leads development of joint attention related actions based on these studies. In chapter 5, therefore, we propose a learning mechanism based on Moore's theory [13].

2.1.3 The difference between individuals

There are many studies on the development of infants with joint attention deficits compared to a normal infants. In particular, understanding development of an infant with autism is an important topic in developmental psychology and neuroscience.

Autism is a severe and pervasive neurodevelopmental disorder characterized by abnormalities in face processing such as avoiding eye-contact with a partner and in social communication skills including joint attention skills [12; 52; 53]. Infants with autism have difficulty in estimating the intentions of others and sharing attention with others. It is, however, suggested that they can acquire some forms of joint attentional behavior such as gaze following and gaze alternation though they acquire

them at an older age than typical infants [54]. Moreover, their development of language and joint attention behavior is promoted when their caregivers often follow their attention and engage in joint attention with them [55]. Behavioral treatment based on operant learning shows the effect of improving communication skills in the infants with autism [56]. This seems to suggest that infants with autism can acquire some forms of joint attentional behavior based on contingency without though their developmental process may be different from that of normal infants.

In this study, we do not build an autism model to compare the difference between the model and a normal infant. In chapter 5, however, we mention that the proposed mechanism is useful to investigate this difference.

2.2 Challenges for Joint Attention in robotics

In robotics, joint attention studies have recently been gaining increasing attention not only from the viewpoint of building communicative robots [57; 58; 59; 60; 61] but also from synthetic approaches to modeling and understanding human developmental processes [16; 62] (see a survey [63]).

Many previous studies have concentrated on the development of gaze following. We can categorize them according to the underlying theory: Moore's theory or Tomasello's. In this section, we review them to clarify the significance of our study by comparison with our study from two viewpoints of development of joint attention and change in social interaction.

2.2.1 Acquisition of gaze following based on learning contingency

Moore and his colleagues [13; 14] found that infants can learn following the gaze of others based on contingent reinforcement. Many previous synthetic studies on the development of gaze following were based on contingency learning. These can be classified in to two categories: with and without explicit evaluation from a caregiver for achieving gaze following.

With external evaluation

Some studies have proposed a learning mechanisms that require evaluation for the robot's behavior from a caregiver. Matsuda and Omori [18] simulated the experiment conducted by Moore and his colleagues and found that a learning agent based on reinforcement learning can acquire gaze following without the concept of understanding the other's intention.

Nagai *et al.* proposed a developmental learning model that involves a robot's development of vision processing, which is represented as a gradual increase of the sharpness of a Gaussian spatial filter for the visual image, and a caregiver's development, which is expressed as adaptive evaluation from a human caregiver according to the robot's performance of gaze following [64]. They indicated that the caregier's help accelerates the robot's learning of gaze following.

Without external evaluation

In naturalistic interaction, a caregiver does not necessarily evaluate an action of an infant. The infant should have an ability to learn gaze following by itself. Therefore, other studies have proposed mechanisms for learning of gaze following without any explicit evaluation from a caregiver. Fasel *at al.* [65] who are members of the MESA (Modeling the Emergence of Shared Attention) project suggested a basic set to enable learning of gaze following without any explicit evaluation from a caregiver. The basic set consists of (1) a set of motivational biases to look at and shift attention between interesting things, (2) a learning mechanism which takes advantage of the temporal structure of predictable, contingent interactions, and (3) a structured environment providing strong correlation between where caregivers look and where interesting things are. They assumed that a shift in the caregiver's gaze implies something salient in the direction of gaze. Based on this assumption, Triesch *et al.* [20] proposed a learning mechanism based on reinforcement learning that enables learning gaze following without any explicit evaluation from a caregiver and showed the effectiveness using computer simulation of face-to-face interaction. Teuscher and Triesch also investigated how the caregiver's behavior influences the learning process of gaze following for infant models with simulated developmental disorders such as autism

and Williams syndrome [66]. As another research, they proposed a model that acquires gaze following developmentally as shown by Butterworth and Jarrett [10] by using depth information from the environment [67] although their previous models did not acquire gaze following in such a way.

These studies were implemented in computer simulation, while our group has proposed mechanisms for a real robot to acquire gaze following. Nagai *et al.* [19] proposed a learning mechanism that enables a robot to acquire gaze following by learning a sensory-motor mapping from the face pattern of a caregiver to its own motor command based on the idea similar to a basic set proposed by Fasel *at al.* [65]. They implemented the proposed mechanism into a real robot, and showed it can acquire gaze following developmentally as shown by Butterworth [10] through interaction with a human caregiver. Our group also proposed models that learn to achieve gaze following for strangers based on generalization through interaction with several caregivers [24] and that acquire gaze following through real-time interaction with a human caregiver within a reasonable period of time by using an automatic attention selector [68]. As another approach, Nagai proposed a model that learns joint attention based on head movement imitation and showed that head movement information accelerate learning of joint attention [69],

These studies showed that a robot can acquire gaze following without understanding the intentions of others. However, they concentrated only on the mechanisms to learn gaze following.

2.2.2 Acquisition of gaze following based on understanding the other's intention

Some studies based on Tomasello's theory have been conducted. While a robot built a direct sensory-motor map from the face pattern of a caregiver to motor commands to achieve gaze following in previous studies based on Moore's theory, they built a controller for visual attention and a predictor about the environment. The predictor learns what to look at through interaction with a caregiver. The controller learns where to look. The predictor is expected to enable a robot to share attention with others.

Hoffman *et al.* [21] proposed a probabilistic model that learns gaze following by imitating a caregiver based on a model for imitation in infants [70]. The model learns a predictor about visual attention of the caregiver through experience of gaze following for a caregiver. However, they did not compare the developmental process of the model with that of a human infant.

Konno and Hashimoto [71] proposed a computational model that enables learning of the gaze alternation to an object outside the visual field. The infant model learns a controller for visual attention and a predictor about what to see at the next time step through experience of looking between an object and a caregiver's face alternately in the visual field. However, it seems to be a quite strong assumption that the infant model always looks between an object and a caregiver's face alternately.

Although these studies are interesting in terms of providing different viewpoints of studies based on Moore's theory, they do not propose models that acquire various forms of joint attentional behavior yet. In addition, their models always need the explicit instruction from caregivers.

2.2.3 Open-ended development in robotics

As described above, previous synthetic studies have proposed models to acquire only one ability of joint attentional behavior. To understand how a robot can share attention with others, we should address how it can acquire various forms of joint attentional behavior. Some researchers begin to study general developmental approaches [72] that may give a clue now to solve such an issue. They have suggested that a robot should be equipped with capacities for autonomous development and with intrinsic motivation systems [63; 72]. Artificial curiosity such as novelty[73] or learning progress [74] has been proposed as intrinsic motivation systems. We cannot, however, evaluate which motivation is appropriate for modeling development of joint attention because the situation where a robot interacts with a partner is not addressed.

2.3 Summary

In this chapter, we gave an overview of the development of joint attention and previous synthetic studies on modeling its development. Human infants seem to acquire various actions related to joint attention in an almost fixed order as shown in Figure 2.1. There still remain unclear points about the underlying mechanism that determines the development order of several forms of joint attentional behavior. However, previous synthetic studies did not focused on the development of several forms of joint attention behavior. In order to realize a mechanism for acquiring several actions without any external instruction from a caregiver, we take account of the fact that infants can not only find social contingency in interaction with a partner but also coordinate their own actions based on the found one.

The next chapter provides a basic environment setting where a robot learns joint attentional behavior and a mathematical explanation about contingency of interaction with a caregiver.

Chapter 3

Interaction for development of joint attention and the contingency

In this chapter, we explain about face-to-face interaction between a caregiver and a robot in this study. Basically, the interaction mentioned here is used in the following chapters 4, 5, and 6. Moreover, we show mathematical explanation about contingency in interaction with a caregiver. We represent the interaction by three types of random variables about sensory information obtained by a robot, the robot's action, and the consequences of its action to evaluate the contingency.

3.1 Face-to-face interaction

Fig. 3.1 shows an environmental setting in this study. A robot sits across from a caregiver at a fixed distance while objects are randomly placed on the table between them. The robot observes environmental information as follows:

1. the robot observes its environment including the caregiver to obtain sensory information $\boldsymbol{S}$ called sensory variables one of which expresses a feature such as gaze direction of the caregiver's face,

2. it also observes the result of last own actions to obtain resultant sensory information $\boldsymbol{R}$ called the resultant sensory variable one of which indicates a kind of an internal reward such as looking at the frontal face of the caregiver, and

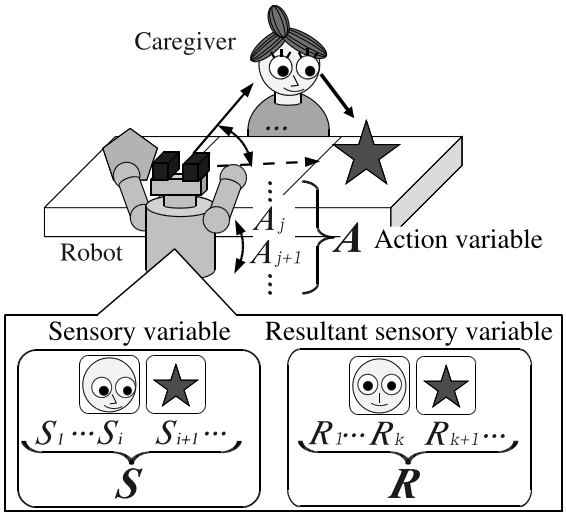


Figure 3.1: Environmental setting for acquisition of joint attentional behavior

3. it takes actions based on motor commands $\boldsymbol{A}$ called action variables one of which indicates a type of an action such as shifting the robot's gaze.

We call a triplet (S_i,A_j,R_k) an event. The robot's task is to find a contingent event from possible ones and to build a sensory-motor mapping from a sensory variable to a motor one based on the selected event to reproduce the contingency.

3.2 Contingency in interaction

We regard the contingency inherent in the interaction as a statistical bias on the state transition in a certain event. The event (S_i,A_j,R_k) has contingency if a robot frequently experiences a state transition in R_k deriving from a pair of states of S_i and A_j. Finding contingency indicates that the robot selects the event that has the strongest bias. Reproducing contingency represents that it performs actions expressing the bias.

We introduce a contingency measure to evaluate a statistical bias in an event, and propose a mechanism to find and reproduce the contingency based on the measure in the following chapters 4 and 5,

Chapter 4

Detection of interaction contingency by transfer entropy

4.1 Introduction

Human infants seem to learn gaze following in uncertain situations in which their caregivers do not always attain joint attention with them. Previous synthetic studies have argued that contingency between gazing behaviors of an infant and its caregiver can be utilized to learn gaze following in such uncertain situations [19; 20]. These studies assumed that a shift in the caregiver's gaze implies something salient in the direction of gaze, and such an object would also be salient to an infant robot.

This assumption implies underlying contingency appearing as statistical bias in infants: they frequently find something salient by looking where a caregiver is looking. Previous studies [19; 20] have shown that a robot can acquire sensory-motor mapping to achieve gaze following by associating a pair of variables involved in such contingent experiences, i.e., the action variable of shifting its gaze and the preceding sensory variable of the direction of the caregiver's gaze. However, no work has, to our knowledge, presented a model to enable a robot to detect such contingency. In other words, how a robot can select contingent pairs of variables from possible candidates has not been addressed. Robots usually have many candidates of variables owing to their multiple perceptual modalities and many degrees of motor freedom. Moreover,

it is unknown what kind of contingency exists in the interaction since modeling human interaction itself is difficult. Building a robot that automatically selects pairs of sensory and action variables that form a contingent structure is therefore formidable.

An important first step in determining this is investigating how contingency in interactions between a caregiver and a robot is quantified. *Transfer entropy* — an information theory measure that detects contingency — appears to be promising in this regard. It shares some of the desired properties of mutual information but also takes into account the dynamics of information transport [23]. Transfer entropy has been shown to need fewer samples and cost in less calculation in detecting causality than other methods for detecting causality such as measures based on Granger causality [75]. Sporns *et al.* showed that a robot with eyes can detect the causal structure inherent in a given sensory-motor coordination, i.e., visual tracking behavior, using transfer entropy [76; 77]. However, they did not address the learning of new behavior based on the found contingency. We studied how transfer entropy can be applied to detect contingency in interactions with a caregiver and how to utilize it to learn new sensory-motor mapping, which appears to be a building block in basic social behavior, i.e., gaze following.

The rest of this chapter is organized as follows. First, we explain the contingency learning reported by Nagai *et al.* [19] as a learning mechanism and the contingency that a robot should find. Next, we introduce a computer-simulated setting involving face-to-face interaction to determine whether transfer entropy enables a robot to find the contingency inherent in interactions with a caregiver. We discuss how to calculate transfer entropy and present experimental results. We show that the robot can acquire gaze following using the found contingency. Finally, discussion on projected issues and concluding remarks are given.

4.2 Contingency detected in gaze following

Figure 4.1 shows gaze following behavior that a robot can acquire based on the learning mechanism proposed by Nagai *et al.* [19]. First, it observes the caregiver's face and then shifts its gaze to follow the caregiver's gaze. Instead of explicitly instructing the robot how to act, they showed that a robot could acquire a sensory-motor map

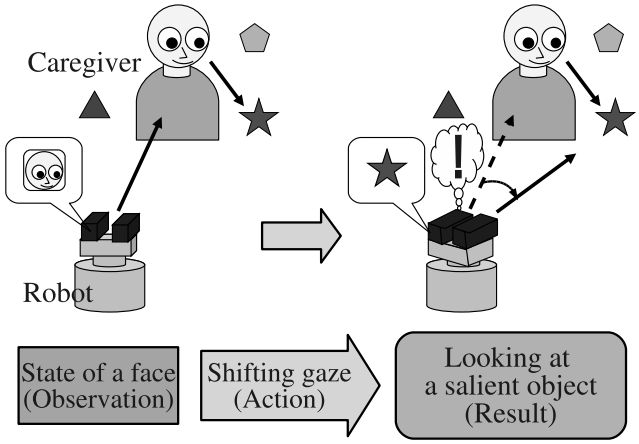


Figure 4.1: gaze following

for gaze following by what they called *contingency learning*.

Since the robot had no experience with gaze following, it sometimes succeeded and sometimes failed to find the same object that the caregiver was looking at. In contingency learning, the robot evaluates only whether it successfully looked at the salient object in both occasionally succeeded and, unfortunately, failed attempts to look at the same object. When it looked at the salient object, its gaze shift and the preceding perception of the caregiver's face pattern (face orientation) were associated. The assumption that the caregiver looks at a salient object for the robot enabled it to acquire gaze following through contingency learning. This tendency derives contingency form its own gaze shift: the robot observed something salient because its gaze frequently followed the direction of the caregiver's gaze. This contingency appears as statistical bias based on frequent experiences of seeing something salient when looking in the direction of the caregiver's gaze. That is, by associating its gaze shift (an action variable) and the caregiver's face direction (a sensory variable), based on the consequences of its action, i.e., the robot observed an salient object, the robot can acquire a sensory-motor map for gaze following.

Nagai *et al.* showed that a robot can acquire gaze following by associating this contingent pair of variables for the consequences of its action even without explicit

instructions. The designers, however, had to specify what kinds of variables should be associated to acquire it. We enhanced contingency learning [19] by investigating whether a robot could automatically find a contingent pair of variables for an the consequences of its action to be associated to acquire gaze following.

4.3 Environmental setting

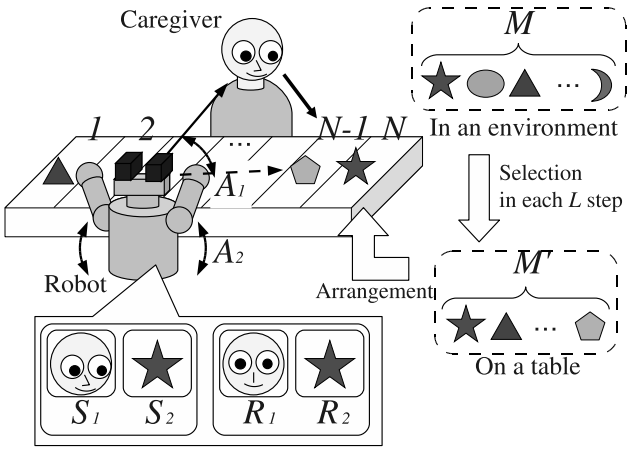


Figure 4.2: Overview of caregiver-robot interaction

Table 4.1: Types of variables in robot

Type	Name	Elements
$\boldsymbol{S}$	caregiver's face	$S_1 = \{f_1, f_2, \cdots, f_N, f_r, f_\phi\}$
	type of object	$S_2 = \{o_1, o_2, \cdots, o_M, o_\phi\}$
$\boldsymbol{A}$	shifting gaze	$A_1 = \{g_1, g_2, \cdots, g_N, g_c\}$
	moving hands	$A_2 = \{h_1, h_2, \cdots, h_{N_h}\}$
$\boldsymbol{R}$	full face of caregiver	$R_1 = \{0, 1\}$
	object	$R_2 = \{0, 1\}$

To determine whether a robot can find a contingent pair of variables for consequent experience in face-to-face interaction with a caregiver, we start with a rough model

of the caregiver's gaze shift. We simulate almost the same interaction as in previous studies [19; 20]; but, for the robot, we add actions such as hand gestures and sensory variables such as types of objects not related to gaze following. This experiment confirms whether the robot can eliminate unrelated variables from candidates for the elements of the sensory-motor map for gaze following.

4.3.1 Environment and interactions between caregiver and robot models

In an experimental computer simulation setting (Figure 5.2), the robot sits across from the caregiver at a fixed distance while objects are randomly placed on the table between them. Let N be the number of positions on the table, M' $(0 < M' < N)$ the number of salient objects placed on spots, and M the number of possible objects. M' objects are selected from M candidates every L steps and spots on which they are placed are determined randomly (only one object per spot). The robot gestures and shifts its gaze, and the caregiver only shifts her gaze. The robot have random variables shown in Table 5.1.

The caregiver and robot take turns observing objects or the other side in each time step as below. First, the caregiver shifts gaze, then the robot observes the caregiver's face or a spot on the table as the current target, obtaining information about S_1, where the caregiver appears to be looking, or S_2, what objects are being observed. We assume that the robot prefers both the caregiver's full face and salient objects to the caregiver's profile because infants appear to prefer the full human face [31] and objects with complex textures or symmetrical patterns [32]. The robot has resultant sensory variables representing such preferences. In the observation timing, it also perceives resultant sensory variables of the caregiver's full face, R_1, and objects, R_2. After observation, the robot gestures (A_2), then shifts its gaze (A_1).

4.3.2 Robot model

Current sensory variable states of the caregiver's gaze, S_1, and objects, S_2, are obtained when the robot observes a target. The direction of the caregiver's gaze in

the t-th step is denoted by $s_1^t \in S_1 = \{f_1, \cdots, f_N, f_r, f_\phi\}$, where $f_1, \cdots, f_N$ indicates at which spot the caregiver is looking, f_r means the caregiver is looking at the robot, and f_ϕ means the robot is not looking at the caregiver's face. The sensory variable for objects in the t-th step indicating what it is looking at is denoted by $s_2^t \in S_2 = \{o_1, o_2, \cdots, o_M, o_\phi\}$, of which $o_1, \cdots, o_M$ correspond to possible objects and o_ϕ indicates that it is looking at something else.

Current states of resultant sensory variables for the caregiver's full face, R_1, and for objects, R_2 are obtained in observation timing. These variables in the t-th step are denoted by $r_1^t \in R_1 = \{0, 1\}$ and $r_2^t \in R_2 = \{0, 1\}$, where "1" means that the robot is looking at its preferred face or an object while "0" means "NOT."

After these observations, it shifts its gaze and gestures. The gaze shift in the t-th step is denoted by $a_1^t \in A_1 = \{g_1, \cdots, g_N, g_c\}$, indicating the target to be gazed at, i.e., a certain location on the table $(g_1, \cdots, g_N)$ or the caregiver's face (g_c). The gesture in the t-th step is denoted by $a_2^t \in A_2 = \{h_1, \cdots, h_{N_h}\}$, indicating the type of movement, and N_h indicating the number of different hand gestures. The robot randomly selects one element in both A_1 and A_2 at each time step.

4.3.3 Caregiver model

A caregiver responds to an infant's behavior and induces the infant's response in interactions with the infant in addition to looking at a salient object as a basic and natural behavior. We modeled behavior so that the caregiver looks randomly at the robot or at one of the objects and shows responsive and inductive behaviors regarding robot behavior.

In the caregiver's gaze shift, three options exist for shifting the gaze when looking at the robot or at an object on the table (Figure 4.3): (1) following the robot's gaze — responding to joint attention (RJA) —; (2) shifting gaze to draw the robot's attention — initiating joint attention (IJA) —; and (3) randomly selecting a target to gaze at (neutral) excluding behavior identical to the RJA and IJA. Note that the caregiver invariably looks at the robot's face or at an object on the table.

In each time step, the caregiver first perceives a target and selects an option based on what is being looked at. If the robot's face is being looked at, the caregiver

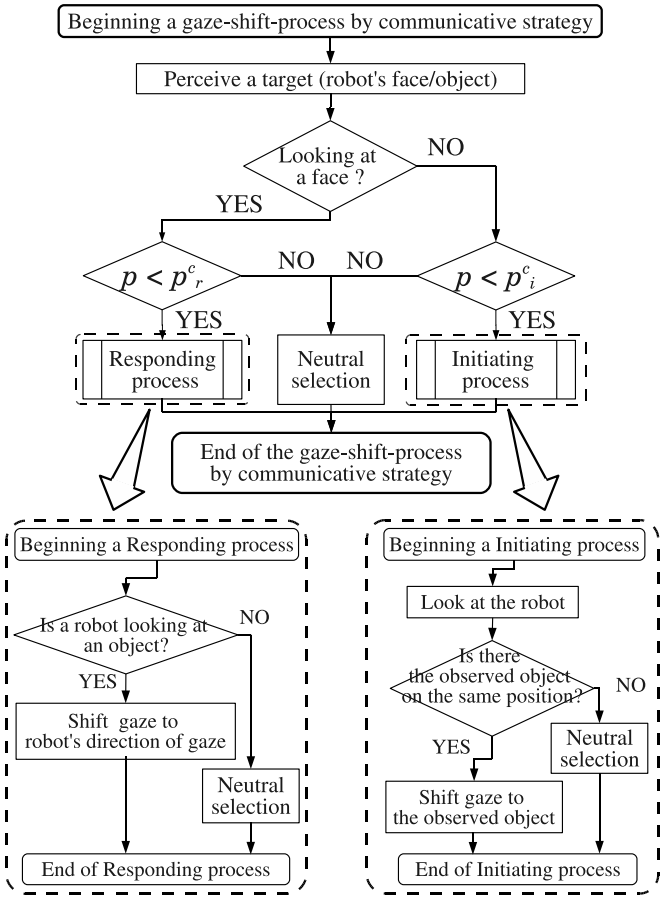


Figure 4.3: Caregiver's gaze shift

selects either RJA with probability p_r^c or the neutral process with probability $1 - p_r^c$. Otherwise, (looking at a object on the table, for example), the caregiver selects either IJA with probability p_i^c or the neutral process with probability $1 - p_i^c$. In RJA, the caregiver shifts her gaze to follow the direction of the robot's face. If the robot is not looking at an object, the caregiver selects an object at random and shifts her gaze to it (box, bottom left, Figure 4.3). In IJA, the caregiver shifts her gaze as if trying to lead the robot's gaze to the object that the caregiver is currently looking at, looking back at the robot and shifting her gaze to the target object in the next step again (box, bottom right, Figure 4.3).

4.4 Transfer entropy

We use transfer entropy [23] to quantify contingency of an event. Transfer entropy is an information measure that represents the flow of information between stochastic variables that cannot be extracted by other information criteria such as mutual information.

We assume that the current state of stochastic variable X is only influenced by the last state of X and the last one of another stochastic variable Y. Transfer entropy that indicates the influence of stochastic variable Y on stochastic variable X is calculated by

$$T_{Y\to X} = \sum_{\substack{x^{t+1},x^t\in X,\\ y^t\in Y}} p(x^{t+1},x^t,y^t)\log\frac{p(x^{t+1}|x^t,y^t)}{p(x^{t+1}|x^t)}, \tag{4.1}$$

where x^t and y^t are observables of X and Y at time step t. This is equivalent to Kullback-Leibler entropy between $p(x^{t+1}|x^t)$ and $p(x^{t+1}|x^t,y^t)$.

We calculate transfer entropy $T_{S_i,A_j\to R_k}$ indicating the influence of a pair of sensory variables S_i $(i=1,2)$ and actions A_j $(j=1,2)$ for resultant sensory information R_k $(k=1,2)$:

$$T_{S_i,A_j\to R_k} = \sum_{R_k,S_i,A_j} p(r_k^{t+1},r_k^t,s_i^t,a_j^t)\log\frac{p(r_k^{t+1}|r_k^t,s_i^t,a_j^t)}{p(r_k^{t+1}|r_k^t)}. \tag{4.2}$$

An observed consequence is often strongly included in contingency inherent in specific actions. Here, shifting the gaze, for example, has a high contingent relationship with the reward for the caregiver's full face: the robot cannot look at the caregiver's face if it shifts its gaze to a spot on the table. In such cases, transfer entropy would not work in finding contingent actions coordinated by any sensory information for reward because the contingency between a result and an action is too strong. We introduce transfer entropy T^C that focuses on the effect of combining sensory and action variables for resultant sensory variable:

$$T^{C}_{S_i,A_j\to R_k} = T_{S_i,A_j\to R_k} - T_{A_j\to R_k}$$
$$= \sum_{R_k,S_i,A_j} p(r_k^{t+1}, r_k^t, s_i^t, a_j^t) \log \frac{p(r_k^{t+1}|r_k^t, s_i^t, a_j^t)}{p(r_k^{t+1}|r_k^t, a_j^t)}, \tag{4.3}$$

which indicates the combinatorial influence of sensory variable S_i and action variable A_j for resultant sensory variable R_k. This appears equivalent to Kullback-Leibler entropy between $p(r_k^{t+1}|r_k^t, s_i^t, a_j^t)$ and $p(r_k^{t+1}|r_k^t, a_j^t)$.

4.5 Experiments

4.5.1 Experimental setup

We conducted a computer simulation to determine whether the robot could find the contingent structure in face-to-face interactions using the proposed contingency measure to acquire gaze following. Calculating transfer entropy requires determining joint probabilities and conditional probabilities for each combination. We estimated them using histograms of observable combinations of three variables: the history of the robot's experience. To demonstrate the potential of the proposed measure, we iterated interaction steps and observed the transition of transfer entropy calculated from histograms.

In experiments, we set nine spots on the table ($N = 9$), ten objects in the environment ($M = 10$), and three objects on the table ($M' = 3$). Note that we set the number of possible objects $M = N + 1$ to nearly equal the number of elements between sensory variables because the finer the resolution of a stochastic variable is, the larger the transfer entropy. For the same reason, the number of hand gestures $N_h = N + 1$. Other parameters $(L, p_r^c, p_i^c) = (10, 0.8, 0.2)$. Experiments lasted while absolute differences between transfer entropies of all combinations of variables between consecutive steps exceed constant value θ. Here, $\theta = 1.0 \times 10^{-7}$.

4.5.2 Transfer entropy in face-to-face interaction

As shown in previous studies [19; 20], the direction of the caregiver's gaze s_t^f leads to a predictable consequence of the robot's shifting its gaze a_t^g, that is, finding a salient object r_2^{t+1}. Conversely, the robot's hand gestures, a_2^t, are not contingent because the caregiver does not respond to them and her gaze direction does not lead to any predictable consequence related to them. We expect the robot to find pair S_1 and A_1 for R_2 matching the pair to which gaze following acquisition is attributed in previous study [19].

Figure 4.4 shows examples of time courses of T^Cs of sensory information, actions, and resultant sensory information in interactions. The vertical axis indicates the logarithmic value of T^C, and the horizontal axis indicates time steps. Since the estimated probability distribution was less accurate at the beginning of interactions, T^Cs seemed overestimated. After interactions are iterated, however, $T^C_{S_1,A_1\to R_2}$ (blue line in Figure 4.4) appeared larger than the others, i.e., the combination of the sensory information of the caregiver's gaze, the change in the robot's gaze, and the reward for salient objects was contingent, indicating that the robot detected a contingent combination of variables with transfer entropy, that was used to acquire gaze following in previous work [19].

To evaluate the robustness of transfer entropy measure for finding a contingent combination of variables, we analyzed the influence of other parameters, such as M', p_r^c, and p_i^c on target transfer entropy ($T^C_{S_1,A_1\to R_2}$) and the difference between target transfer entropy and the highest transfer entropy among other combinations ($\max_{k,l,m} T^C_{S_i,A_j\to R_k}$). Note that this difference must be larger than zero for target combinations of variables to be contingent. We call the difference ΔT^C_{diff}. We varied p_r^c, p_i^c, and M' at 0.25, 0.50, and 0.75 for p_r and p_i, and $M' = 1, 2, \cdots, 9$. For each parameter setting, we ran ten 30,000-step simulations and plotted the averages and standard deviations of ΔT^C_{diff} in the 30,000-th step for the number of objects in Figure 4.5. Note that ΔT^C_{diff}s for most parameter settings exceeded zero except in the case of $M' = 8, 9$, confirming that the target combination of variables was contingent for all parameter settings except extreme cases in which almost all places are salient for the robot, although absolute differences appear to reflect the number

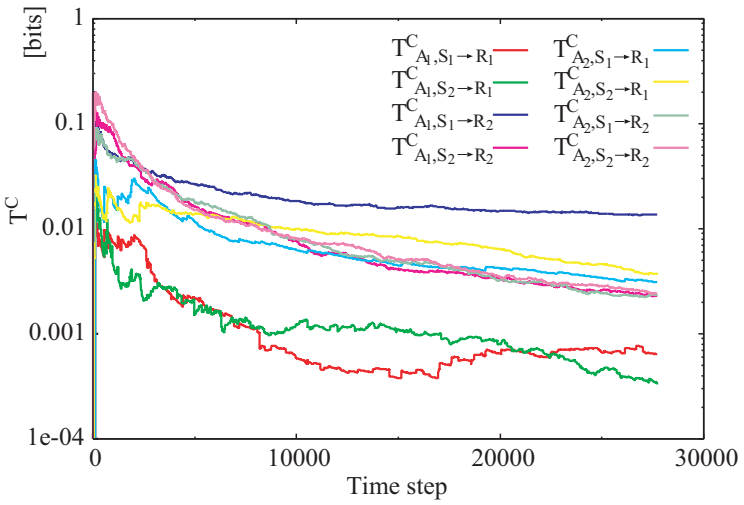


Figure 4.4: Time courses of contingent measure of combinations of variables in face-to-face interactions between caregiver and robot

of objects M'. Note also that p_r^c and p_i^c do not affect ΔT^C_{diff} from standard deviations in Figure 4.5.

4.5.3 Influence of uncertain contingency

In actual interaction between a caregiver and infant, the caregiver may look at an object not salient to the infant. Therefore, we examined to what extent the proposed mechanism depends on the assumption that a caregiver tends to look at something salient to the infant.

We changed the caregiver model to one that behaves as described in Section 4.3-C with probability p_c^c and looks at the robot or an empty spot with probability $1 - p_c^c$. If we set p_c^c to a lower value, the caregiver looks less often at an object and more often at empty spots on the table and behaves completely randomly around $p_c^c = 0.5$. We compared the transfer entropies calculated in the 30,000-th step in interactions with different of p_c^c under the above setting.

Figure 4.6 shows the averages and standard deviations for ten simulations of ΔT^C_{diff}. Since the difference became positive and $T^C_{S_1,A_1\to R_2}$ had a higher value when p_c^c exceeded 0.6, the proposed mechanism appeared effective when the caregiver looked

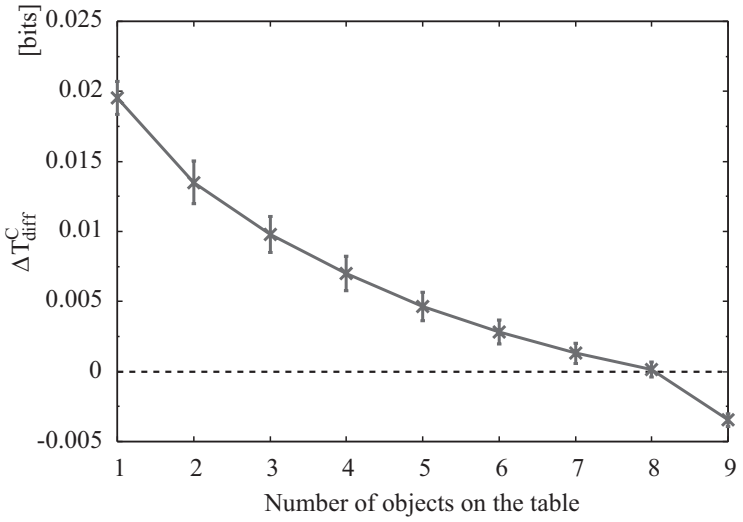


Figure 4.5: Change of difference between $T^C_{S_1,A_1 \to R_2}$ and largest T^C in other combinations in situations of different combinations of p^c_i, p^c_r, and M'

sometimes at objects salient to the robot. Note that the difference again became positive when $p^c_c < 0.2$, meaning that T^C detects opposite contingency, i.e., if the robot follows the direction of the caregiver's gaze, it cannot look at any salient objects. The proposed mechanism detects contingent combinations in face-to-face interaction regardless of whether structures are related to the acquisition of gaze following. The robot thus use the detected combination to acquire gaze following if the caregiver looks often at objects salient to the robot, i.e., $p^c_c > 0.6$.

4.5.4 Learning gaze following with detected contingent variables

We studied whether a combination of variables with maximum T^C ($T^C_{S_1,A_1 \to R_2}$) enables the robot to learn gaze following. Before having the robot do so, we confirmed that contingency of found combinations showed the robot's experience from which it learned gaze following. Figure 4.7 shows histograms of experience in which the robot observed the caregiver's face and chose to shift its gaze to a spot before observing an object through interaction. Diagonal elements correspond to gaze following, and

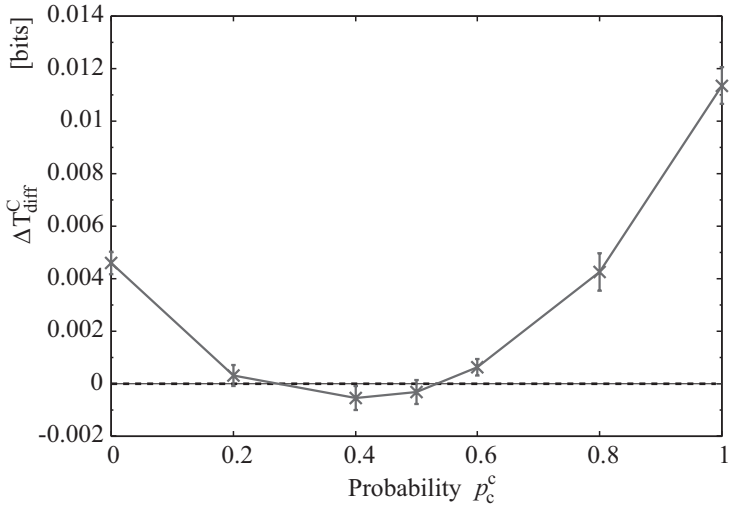


Figure 4.6: Change in difference between $T^C_{S_1,A_1 \to R_2}$ and highest T^C in other combinations based on probability p^c_c

showed that the robot tended to successfully observe an object when it occasionally performed the same behavior as gaze following.

As shown by Nagai *et al.*, a robot acquires gaze following using contingency learning [19] in situations where the robot's experience is biased to occasionally achieve successful gaze following. In subsequent computer simulation, we examined whether it obtained a sensory-motor map for gaze following by contingency learning based on the detected pair of sensory and action variables.

Sensory and action variables included in the contingent combination with the highest T^C were assigned to input and output layers of a two-layered perceptron (Figure 4.8). Since contingency learning was conducted by associating sensory-motor variables regardless of gaze following success, it is implemented using the current observable action variable A_1 as the desired value of the output layer in backpropagation learning. The perceptron was trained with data obtained through 30,000 interactions in which actions of the caregiver and the robot were determined by models described in Section 4.3. Sensory (action) variables were encoded so that input (output) to only one node was "1" while others "0". Suppose that the robot finds something salient ($r_2^{t+1} = 1$) by shifting its gaze to the i-th spot on the table ($a_1^t = g_i$) after observing

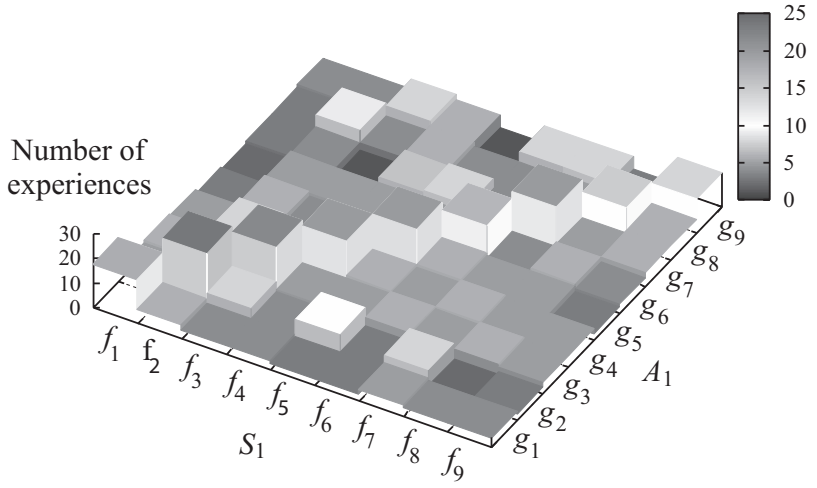


Figure 4.7: Distribution of experiences to $R_2 = 1$ in interactions between a caregiver and a robot. In this graph, pairs with the identical number of suffixes, e.g., (f_1, g_1), correspond to behavior of gaze following

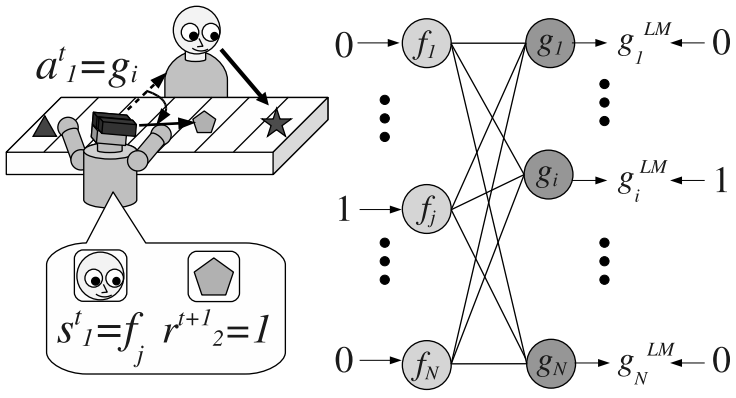


Figure 4.8: Network to learn gaze following.

the caregiver's face, perceived as looking in the j-th direction, ($s_1^t = f_j$). The perceptron receives an input vector of which f_j is one while the others are zeros and receives a reference vector of which g_i is one while the others are zeros (Figure 4.8).

After ten trials, each consisting of 30,000 interactions, we examined the average success rate for gaze following, testing whether the perceptron output the action variable corresponding to the caregiver's gaze for each of N perceptual inputs. Success rate for each trial was calculated by the number of pairs of input and output achieving gaze following. Average success rate was 84%, confirming that contingent variables selected by the proposed mechanism can be utilized to learn gaze following.

4.6 Summary and discussion

In this chapter, we showed that transfer entropy is promising in detecting the contingency inherent in face-to-face interaction. Transfer entropy helps a robot detect important variables constituting the contingent structure inherent in interaction. We also demonstrated that appropriately chosen contingent variables can be used in learning gaze following.

Influence of other parameters We did not focus on parameters L, N, M, and N_h in experiments because the behavior of L and N is easily predicted. As either L or N increases, a robot needs more interactions to detect the target combination (S_1, A_1, R_2) because transfer entropies are overestimated due to the inaccuracy of the estimated probability distribution. We set M and N_h as $N + 1$ to reduce differences between transfer entropies that attribute to different numbers of possible elements of variables. An infant, however, appears to have the different resolution of multimodal sensations and various kinds of actions because these components develop in parallel and at different time schedules. We should therefore utilize normalized transfer entropies in the number of elements to adequately estimate the contingency of combinations that consist of different numbers of elements.

Mutual response Experimental results showed that responsive and inductive behaviors of a caregiver influence the contingency inherent in interactions between the

caregiver and a robot only negligibly because the robot did not respond to the caregiver's actions. The caregiver's behavior helps the robot to detect the contingent combination for gaze following if we design the appropriate robot responses to the caregiver's actions. We should also add other action modalities, such as pointing or vocalization to the caregiver. We plan to study what sort of contingency is detected in interaction with such mutual responses.

Improvement toward biologically plausible mechanism Observations in developmental psychology imply that many contingent structures are inherent in infant-caregiver interaction [32]. Infants start to become sensitive to social contingency from about three months of age [45], and acquire related social skills [47; 38]. Such a contingent structure is used to acquire gaze following [14]. Our mechanism appears plausible in that a robot acquires gaze following only by finding the contingency of interactions with humans. We cannot yet, however, explain information processing in the human brain for detecting such contingency. We plan to use mechanisms to detect such contingency in the human brain to propose biologically plausible mechanisms.

Application to other social skill Our mechanism can be applied to the acquisition of other social skills besides gaze following. As stated by Triesch *et al.* [20], the acquisition of point following, defined as looking at an object that someone else is pointing at, appears based on a contingency similar to gaze following, the contingency between the infant's gaze shift and the caregiver's hand use when looking at a salient object. Our mechanism may also enable social skills to be cumulatively acquired. If a robot acquires and use new behavior, this behavior changes the caregiver's behavior and modifies contingent structure, leading the robot to acquire subsequent behavior. Through such acquisition, we expect the mechanism to help us understand what sorts of relationships should be found between the developmental processes of skills and how a caregiver should behave to help a robot acquire skills more easily.

Adaptability to environmental change We expect that skills acquired by a robot will be suitable to individual humans and tasks. Useful social skills are required by social robots to communicate smoothly with humans. Pre-programming such

abilities is, however, difficult because the usefulness of social skills depends on whom the robots communicates with and what tasks they are involved in. As one key to avoid such difficulty, we focused on the fact that many social skills are contingent in interactions with humans. We expect that our mechanism will help us realize social robots with social skills appropriate to humans and tasks.

Evaluating performance of the proposed measure Although the proposed measure seems useful to find contingency in face-to-face interaction with a caregiver, we do not compare it with other measures [75] such as Granger causality and mutual information and examine what feature it has yet. It is beneficial to make clear the limitation of the measure because of its application to other issues. Therefore, we should analyze the features of the proposed measure as a future work.

Chapter 5

Reproducing social contingency toward open-ended development of Joint Attention

5.1 Introduction

Human infants acquire a variety of social actions and gradually attain smooth communication with others. In particular, the ability to achieve joint attention with others is the basis to share attention among agents since the direction of the gaze of a person often indicates where the person's attention is being directed. Therefore, understanding how infants develop actions related to joint attention such as gaze following, pointing, gaze alternation, and social referencing is a central topic in developmental psychology [7]. Infants seem to acquire various kinds of actions gradually in the development of joint attention; after learning gaze following, they begin to show gaze alternation, i.e., successive looking between a caregiver and an object, social referencing, and pointing [11]. However, it is still a mystery why most infants acquire several forms of joint attention behavior in such order.

In robotics, joint attention studies have been done not only from the viewpoint of building communicative robots [57] but also from synthetic approaches to modeling and understanding human developmental processes [16]. Previous synthetic studies

have addressed how infants can acquire gaze following without explicit instructions about where to look [19; 20]. A contingent structure has been shown in a sequence of gazing actions of an infant robot and its caregiver that enables it to successfully acquire gaze following based on statistical association of these actions. The robot acquired gaze following by associating the direction of the caregiver's gaze and own gazing actions, in other words, reproducing the contingency. However, in these studies, the robot was given *a priori* knowledge about which pair of sensory and action variables should be associated. Communicative robots usually have many candidates for sensory and action variables to be associated to acquire such social actions because they are supposed to have multimodal sensori-motor experiences that reflect the contingency in interaction with humans. This indicates that it is not trivial for a robot to select such a pair of sensory and action variables by itself to model contingencies involved in the interaction.

In order to select appropriate variables among many candidates, the learning mechanisms based on intrinsic motivation such as novelty [73] or learning progress [74] seem useful because they enable a robot to select actions to be acquired without designing task-dependent reward. Although these intrinsic motivations enabled robots to acquire several kinds of behaviors, the effects of taking the acquired social actions on interaction with a caregiver were not handled. If a robot prefers to contingency inherent in interaction with a caregiver, it can find the contingency by itself and acquire a social action by reproducing the found contingency as shown in previous studies [19; 20]. We hypothesize that reproducing the found contingency further leads novel contingency to emerge from interaction with a caregiver because it introduces the change of the caregiver's response to the robot into the interaction. We expect this loop of finding and reproducing the contingency to enables open-ended development of social actions. In addition, such preference to the contingency in interaction with a caregiver seems reasonable as a model of an intrinsic motivation in human infants because developmental psychologists observe that they are sensitive to the contingency called contingency [22].

Therefore, we focus on contingencies in pairs of sensory and motor variables to select a pair involving contingency to be reproduced. To determine to what extent

variables constitute strong causality [76; 77; 78], we use the transfer entropy measure [23]. This information theoretic measure shares some of the desired properties of mutual information but also considers the dynamics of information transport. In computer simulations of face-to-face interaction between a robot and a caregiver agent, transfer entropy was confirmed to be useful for the robot to find an appropriate combination of variables that enables to learn gaze following [79]. Such a measure of contingency is expected useful for acquiring not only gaze following but also different kinds of joint attention behavior. We expect to model the developmental process of joint attention by finding contingency and its reproduction: the joint attention behavior acquired by a robot will change the caregiver's response and induce novel contingency in the interaction to acquire another action related to joint attention.

In this chapter, we propose a mechanism that iteratively acquires social actions by extending the measure proposed in our previous work [79]. The mechanism not only finds a combination of contingent variables and constructs a sensori-motor mapping to reproduce behavior based on the found contingency but also selects which sensori-motor mapping should be used to select the next action. In the iterative process, a new variable expressing whether each sensori-motor mapping is used is added to promote finding the contingency depending on other contingencies. In computer simulations, we examine what kinds of actions related to joint attention can be acquired in order by changing the actions of the caregiver agent. The results indicate that a robot acquires a series of actions related to joint attention in order that almost matches with an infant's development of joint attention. The difference between them is discussed based on the analysis of the robot behavior and future issues are given.

The rest of the chapter is organized as follows. We give the learning mechanism for sequential acquisition of social actions. Next, we introduce a computer-simulated setting involving face-to-face interaction to determine what actions a robot acquires through such interaction and then show our experimental results with it. Finally, discussion on future issues and the results of the analysis of the robot behavior and concluding remarks are given.

5.2 Proposed mechanism to successively develop social behaviors

Figure 5.1 shows a mechanism that enables a robot to acquire social actions based on finding contingency inherent in interaction and reproducing the found one. The mechanism consists of four modules: (1) a contingency detector, (2) contingency reproduction modules (CMs), (3) reactive behavior modules (RMs), and (4) a module selector. The number of RMs is constant, while there are no CMs at beginning of learning. They are generated by the contingency detector once it finds the contingency through interactions between a caregiver and the robot.

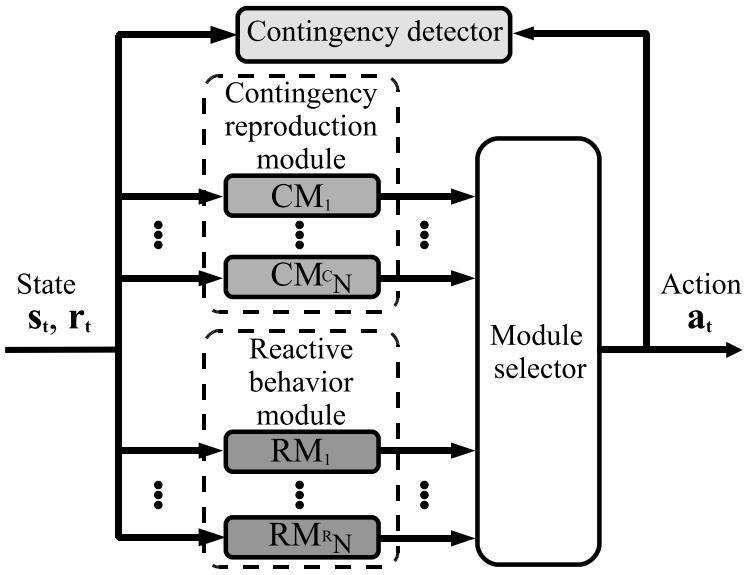


Figure 5.1: Proposed mechanism to successively develop social actions

RMs and CMs output not only motor commands to be executed but also values of their reliability for the current state $\boldsymbol{r}, \boldsymbol{s}$. The reliability indicates how much a motor command selected by each of RMs and CMs is appropriate for the current state in terms of information theory. The module selector determines an action $\boldsymbol{m}$ of the robot based on the reliability. The state and the selected commands are used by the contingency detector to constitute a new CM as well as to determine a contingent

event based on information theoretic measure described in the following section.

5.2.1 Contingency detector

A contingency detector has two main roles: finding a contingent event and generating a new CM based on it. We proposed an information theoretic measure of contingency based on transfer entropy [23] to quantify the contingency inherent in events experienced through interactions with a caregiver. Transfer entropy is a kind of information measure that represents the flow of information between stochastic variables that cannot be extracted by other information criteria such as mutual information. The contingency detector evaluates contingency in interaction by calculating the measures for all events.

Here, we assume that the current state of stochastic variable X is only influenced by the last states of X and another stochastic variable Y. Transfer entropy that indicates the influence of Y on X is defined by

$$T_{Y\to X} = \sum_{\substack{x^{t+1},x^t\in X,\\ y^t\in Y}} p(x^{t+1},x^t,y^t)\log\frac{p(x^{t+1}|x^t,y^t)}{p(x^{t+1}|x^t)}, \tag{5.1}$$

where x^t and y^t are the observables of X and Y at time step t, respectively.

Suppose that combinatorial joint probabilities are given for all possible events. To quantify joint effect of sensory variable S_i and action variable A_j on resultant sensory variable R_k, we introduce state-action contingency (SAC) $C^j_{i,k}$, which is defined and expanded as follows:

$$\begin{aligned} C^j_{i,k} &= T_{(S_i,A_j)\to R_k} - (T_{S_i\to R_k} + T_{A_j\to R_k}) \\ &= \sum_{\substack{s_i^t\in S_i,\\ r_k^t\in R_k}} p(r_k^t,s_i^t) \sum_{\substack{r_k^{t+1}\in R_k,\\ a_j^t\in A_j}} e(r_k^{t+1},a_j^t;r_k^t,s_i^t), \end{aligned} \tag{5.2}$$

where $e(r_k^{t+1},a_j^t;r_k^t,s_i^t)$ in Eq. (5.2) is called a contingent saliency under a pair of observables (r_k^t,s_i^t).

$$e(r_k^{t+1},a_j^t;r_k^t,s_i^t) = p(r_k^{t+1},a_j^t|r_k^t,s_i^t)\log\frac{p(r_k^{t+1}|r_k^t,s_i^t,a_j^t)}{p(r_k^{t+1}|r_k^t,s_i^t)} - p(r_k^{t+1},a_j^t|r_k^t)\log\frac{p(r_k^{t+1}|r_k^t,a_j^t)}{p(r_k^{t+1}|r_k^t)} \tag{5.3}$$

$e(r_k^{t+1}, a_j^t; r_k^t, s_i^t)$ represents a statistical bias on the state transition from r_k^t to r_k^{t+1} deriving from a pair (s_i^t, a_j^t). If the resultant experience r_k^{t+1} depends on a triplet (s_i^t, r_k^t, a_j^t), the difference between $p(r_k^{t+1}|r_k^t, s_i^t, a_j^t)$ and $p(r_k^{t+1}|r_k^t, s_i^t)$ in the first term of Eq. (5.3) becomes larger. However, there is a possibility that the difference derives from a dependency only on a pair (a_j^t, r_j^t). Therefore, the second term in Eq. (5.3), which represents influence of only a_j^t on the state transition from r_j^t to r_j^{t+1}, is subtracted from the first term to capture a combinatorial bias in an event.

After calculating the SACs for all triplets, the detector evaluates whether to generate a new CM for an event with the highest SAC value. Here, a new CM for a SAC is generated if its SAC keeps the highest value during T^C time steps and the absolute difference between the highest SAC and the second highest one between the last consecutive steps is lower than a constant value θ. Hereafter, a CM that is constituted for an event of S_i, A_j, and R_k is denoted as CM(i, j, k).

When the contingency detector creates the i-th new CM, the set of events is extended by adding new variables related to the CM, A_{Π_i} and S_{Π_i}. Here, A_{Π_i} represents whether output from the i-th CM is used as a current motor command to perform its current action while S_{Π_i} represents whether output from the i-th CM was used to perform the previous action at the last step. Therefore, after the number of generated CMs is N^{Π}, the contingency detector calculates the SACs $C_{i,k}^j$ where $S_i \in \{S_1, \cdots, S_{N^S}, S_{\Pi_1}, \cdots, S_{\Pi_{N^\Pi}}\}$ and $A_j \in \{A_1, \cdots, A_{N^A}, A_{\Pi_1}, \cdots, A_{\Pi_{N^\Pi}}\}$, and N^S and N^A indicates the numbers of pre-determined sensory variables and motor variables, respectively.

5.2.2 Contingency reproduction module

A CM is composed of a sensori-motor map from a sensory variable S_i to a action variable A_j of the found event (S_i, A_j, R_k). The map is built to output the contingent motor command under each pair of observables. Here, the contingent motor command is defined as the motor command with the highest contingent saliency of all ones under a pair of observables because a contingent saliency under a pair of observables represents effectiveness of an action in interaction. Therefore, the contingent motor command a_j^* and the expected resultant sensory information r_k^* under a pair

of observables (r_k, s_i) are given by:

$$(r_k^*, a_j^*) = \underset{r_k', a_j'}{\operatorname{argmax}}\, e(r_k', a_j'; r_k, s_i). \tag{5.4}$$

A CM also calculates the reliability for the contingent motor command under a pair of observables. This measure is used by a module selector as described in the section 5.2.4 . We define the reliability $Z(r_k, s_i)$ as z-score for the highest contingent saliency under a pair of observables:

$$Z(r_k, s_i) = \frac{e_{max}(r_k, s_i) - \mu^{s_i, r_k}}{\sigma^{s_i, r_k}}, \tag{5.5}$$

where $e_{max}(r_k, s_i) = e(r_k^*, a_j^*; r_k, s_i)$, and μ^{r_k, s_i} and σ^{r_k, s_i} denote the average of the contingent saliencies under observables (r_k, s_i) and the standard deviation, respectively.

However, not every pair of observables will necessarily be involved in the contingency found in the interaction. The contingent saliencies under the pairs of observables involved in the contingency should be higher than those uninvolved in the contingency. In addition, the differences between the highest contingent saliency and the others under a pair of observables involved in the contingency should also be large. Therefore, we evaluate the reliability as follows:

1. we calculate the average of all the positive or null contingent saliencies (ignoring the negative ones) under all pairs of observables;

2. we check if the highest contingent saliency under each pair of observables exceeds the average. If this is not the case, no further computation is performed;

3. we determine the z-scores of the highest contingent saliency and the second highest one under each pair of observables that satisfies the condition of step 2; and

4. if the difference between the z-scores of the highest contingent saliency and the second one is higher than Z_θ, the reliability under the pair is determined, otherwise the reliability is not computed.

5.2.3 Reactive behavior module

A RM outputs a motor command based on the behavior policies pre-programmed by a designer. Here, we use random selection. We could also use more biased selection because infants seem to have innate preferences, such as preferences to human faces or objects with complex textures. A RM outputs a constant value α as the reliability under each pair of observables too.

5.2.4 Module selector

As the number of CMs increases, a robot must determine which outputs from the CMs and RMs to be selected. A module selector serves this purpose. The module selector determines an action of a robot from outputs of the CMs and RMs based on values of their reliability.

Let ${}^{C}N^{i}$ and ${}^{R}N^{i}$ denote the numbers of CMs and RMs that output motor commands for the i-th joint of a robot, respectively. Here, we express ${}^{C}N^{i}$ CMs and ${}^{R}N^{i}$ RMs as $\Gamma^{i}_{j} \in \{\mathrm{RM}^{i}_{1}, \mathrm{RM}^{i}_{2}, \cdots, \mathrm{RM}^{i}_{{}^{R}N^{i}}, \mathrm{CM}^{i}_{1}, \mathrm{CM}^{i}_{2}, \cdots, \mathrm{CM}^{i}_{{}^{C}N^{i}}\}$. Probability $\Pr(\Gamma^{i}_{j})$ of choosing the output from Γ^{i}_{j} at the t-th step is given by softmax selection such as:

$$\Pr(\Gamma^{i}_{j}) = \frac{\exp(Z^{j}(r^{t}, s^{t}, t')/\tau_{i})}{\sum_{l \in {}^{R}N^{i} + {}^{C}N^{i}} \exp(Z^{l}(r^{t}, s^{t}, t')/\tau_{i})}, \tag{5.6}$$

where $Z^{l}(s^{t}, r^{t}, t')$ is introduced to avoid producing the same behavior continuously such as keeping fixation on the same target. It indicates the value of reliability of the l-th CM that continues to receive the same observables during last t' steps: $Z^{l}(r^{t}, s^{t}, t') = Z^{l}(r^{t}, s^{t})e^{-\beta t'}$. The parameter β is a positive constant and τ_{i} is the temperature parameter. If τ_{i} is set as a constant value, the increase of ${}^{C}N^{i}$ decreases the $\Pr(\Gamma^{i}_{j})$. So, we decrease τ_{i} as ${}^{C}N^{i}$ increases: $\tau_{i} = 1/({}^{C}N^{i} + 1)$

5.2.5 Sequential acquisition of behavior based on reproducing the acquired behavior

At the beginning of learning, the module selector selects the outputs of RMs that output pre-programmed actions as current motor commands of a robot. As interaction between a caregiver and the robot is iterated, however, the contingency detector

selects a contingent event and generates a new CM that constructs a sensori-motor map based on the found event. The CM outputs a contingent motor command for each pair of observables to reproduce the contingent relation in the event. The contingency detector adds S_Π and A_Π as sensory and action variables involving the CM, and comes to evaluate events including them, too. Through the loop of finding the contingent event and reproducing the contingent motor commands, a new CM for an event involving S_Π or A_Π can be generated. As a result, the robot acquires actions that are related to each other.

5.3 Computer simulation of developing joint attention related behavior

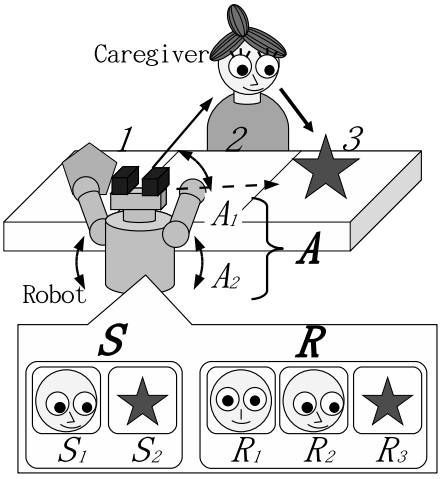


Figure 5.2: Experimental setting for acquisition of actions related to joint attention

The performance of the proposed model was tested in computer simulations where an infant model (hereafter a robot) interacted with a caregiver model in face-to-face situations. The robot selected actions based on the proposed mechanism and caregiver actions were simulated more faithfully as described in the section 5.3.1.

5.3.1 Experimental setting

Environment and Infant model

Figure 5.2 shows an overview of the setting in the computer simulation. There are three spots on a table, and two objects are randomly placed. The spots on which they are placed are determined randomly every ten steps (no more than one object at one spot). Here, we assume that the caregiver and the robot take turns observing environmental information.

Table 5.1: Initial variables in robot

Type	Name	Variables
$\boldsymbol{S}$	caregiver's face	$S_1 = \{f_1, f_2, f_3, f_r, f_\phi\}$
	object	$S_2 = \{o, o_\phi\}$
$\boldsymbol{A}$	shifting gaze	$A_1 = \{g_1, g_2, g_3, g_c\}$
	hand gesture	$A_2 = \{h_1, h_2, h_3, h_4\}$
$\boldsymbol{R}$	frontal face of caregiver	$R_1 = \{0, 1\}$
	caregiver profile	$R_2 = \{0, 1\}$
	object	$R_3 = \{0, 1\}$

The variables in Table 5.1 were set as initial ones in the robot's program. The direction of the caregiver's gaze is denoted by S_1 each member of which indicates the gaze to a particular location of a table (f_1, f_2, f_3) or the robot' frontal face (f_r), or indicates that the robot is not looking at the caregiver's face (f_ϕ). The sensory variable for objects representing whether the robot is looking at an object is denoted by S_2 each member of which indicates that the robot is looking at an object (o) or something else (o_ϕ). We prepare three types of variables as resultant sensory variable: caregiver's frontal face R_1, the caregiver's profile R_2, and objects R_3. These are binary variables indicating whether the robot is looking at its preferred face or an object ("1") or not ("0").

The robot shifts its gaze and gestures. The robot's gaze shift is denoted by A_1 each member of which indicates the target to be gazed at, i.e., a particular location on the table (g_1, g_2, g_3) or the caregiver's face (g_c). The gesture is denoted by A_2 each member of which indicates the four different hand gestures. Here, parameters about

the proposed mechanism were set as $(T^C, \theta, Z_\theta, \alpha, \beta) = (30, 5.0 \times 10^{-5}, 0.5, 1.0, 0.5)$.

The behavior rules of the caregiver model

The robot moves its hands and shifts its gaze, while the caregiver only shifts the gaze. In the chapter 4, we modeled the caregiver's behavior so that the caregiver not only randomly looks at the robot or at one of the objects but also shows responsive and inductive behaviors. Here, we adopt a similar model for the caregiver, except that the current model shows responsive behavior also when it achieves joint attention with the robot.

The caregiver always looks at the robot's face or an object on the table. In the caregiver's gaze shift, three options exist for shifting the gaze when looking at the robot or at an object on the table: 1) following the robot's gaze (RJA process); 2) shifting gaze to draw the robot's attention (IJA process); and 3) randomly selecting a target to gaze at (neutral process) excluding behavior identical to the RJA and IJA processes.

In each time step, the caregiver selects an option based on what she is looking at. The caregiver basically selects neutral process and selects the RJA or IJA processes in the following cases:

- if the caregiver is looking at the robot's face directed to a spot on the table, the RJA process is selected with probability p_r^c and the neutral process is selected otherwise; and

- if the caregiver is looking at an object, the IJA process is selected with probability p_i^c and the neutral process is selected otherwise.

In addition, the caregiver shifts the gaze to the robot's face with probability p_e^c if the caregiver and the robot successfully look at the same object regardless of the selected option. In the RJA process, the caregiver shifts the gaze to follow the direction of the robot's face. If the robot is not looking at an object, the caregiver selects an object at random and shifts the gaze to it. In the IJA process, she shifts the gaze as if trying to lead the robot's gaze to an object that she is currently looking at. She looks back at the robot and shifts her gaze to the target object in the next step again.

5.3.2 Sequential acquisition of joint attention behavior

We ran ten 100,000 steps simulations where the parameters were set as $(p_r^c, p_i^c, p_e^c) = (0.5, 0.5, 1.0)$. The average number of CMs found by the contingency detector was 3.5. In 90 % of the simulations, a particular set of CMs was generated in a fixed order, which was CM$(1, 1, 3)$, CM$(\Pi_1, 1, 1)$, and then CM$(\Pi_1, 1, 2)$. Moreover, they were often generated earlier than other CMs for different events. Each of these CMs enabled a robot to achieve social behavior: following a caregiver's gaze (CM$(1, 1, 3)$; hereafter **FG-m**), shifting its gaze to the caregiver after gaze following for the caregiver (CM$(\Pi_1, 1, 1)$; hereafter **SCf-m**), and shifting its gaze to the caregiver regardless of gaze following (CM$(\Pi_i, 1, 2)$; hereafter **SC-m**). Moreover, they were often generated earlier than other CMs for different events.

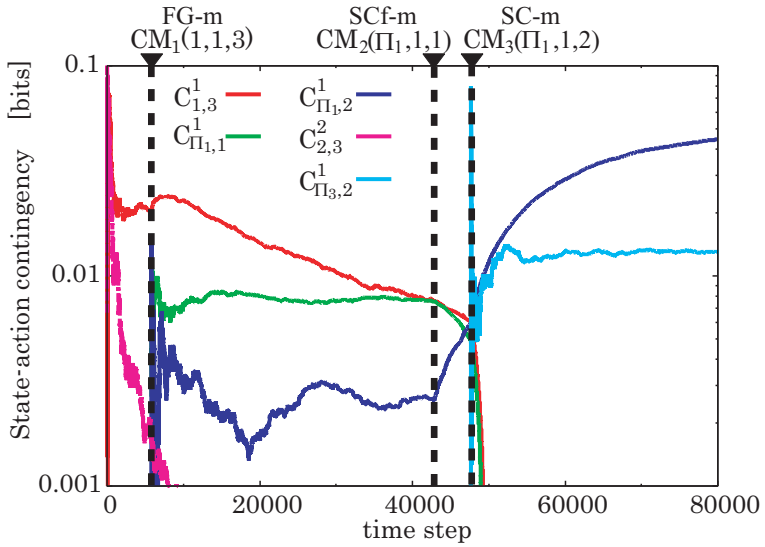


Figure 5.3: Time courses of the state-action contingency of events in a simulation face-to-face interactions between a caregiver and a robot.

Figure 5.3 shows examples of time courses of SACs for several events which have ever been one of the two highest SAC values through 80,000 steps in a simulation. The vertical axis indicates the logarithmic value of the SACs. We also show the timing of generating new CMs as inverted black triangle on the top of the graph. After interactions were iterated, $C^1_{1,3}$ first became the highest among all SACs (red

curve in Figure 5.3). As a result, the FG-m was generated at the 5748-th step, and S_{Π_1} and A_{Π_1} were added as sensory and action variables, respectively. The robot then began to follow the caregiver's gaze by using output from FG-m when it looked at the caregiver who was looking at an object. Through iterating the interaction, $C^1_{1,3}$ gradually decreased because using particular output based on acquired sensori-motor map makes the difference between $p(r_3^{t+1}|r_3^t, s_1^t, a_1^t)$ and $p(r_3^{t+1}|r_3^t, s_1^t)$ (the first term of Eq. (5.3)) smaller. The decrease made $C^1_{\Pi_1,1}$ the next highest value, and the SCf-m whose sensory variable S_{Π_1} is related to using output from FG-m was generated at the 42862-th step. It enabled the robot to direct its gaze to the caregiver after gaze following for the caregiver,

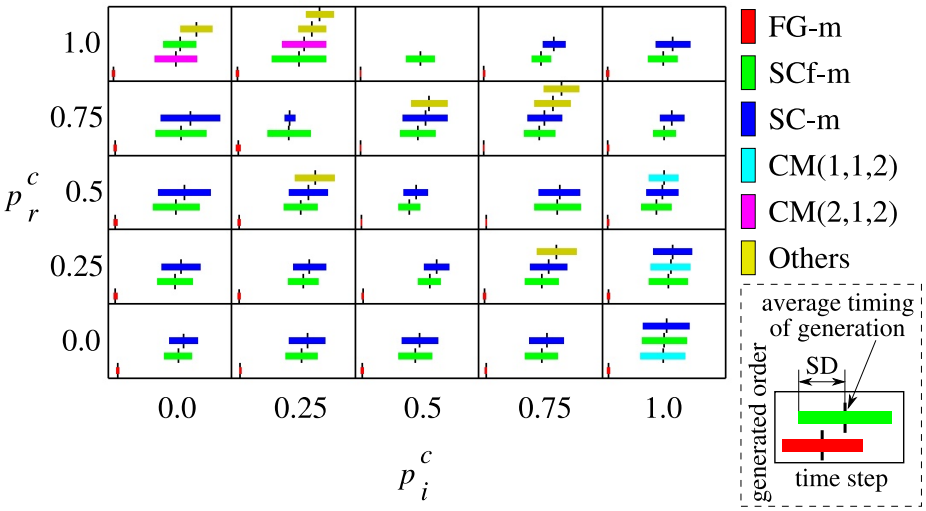
Using output from SCf-m changed the contingency in interaction again and promoted increase of $C^1_{\Pi_1,2}$ (blue curve in Figure 5.3). This caused the generation of SC-m for the event (S_{Π_1}, A_1, R_2) at the 47686-th step. It enabled the robot to shift its gaze to the caregiver despite achieving following the caregiver's gaze. As a result, the robot alternately shifted its gaze between a caregiver and an object, that is, it acquired gaze alternation. This indicates that a robot acquired not only gaze following but also gaze alternation through the repetition of finding and reproducing a chain of contingencies in interaction that change by using output from existing CMs.

5.3.3 Influence of caregiver's behavior

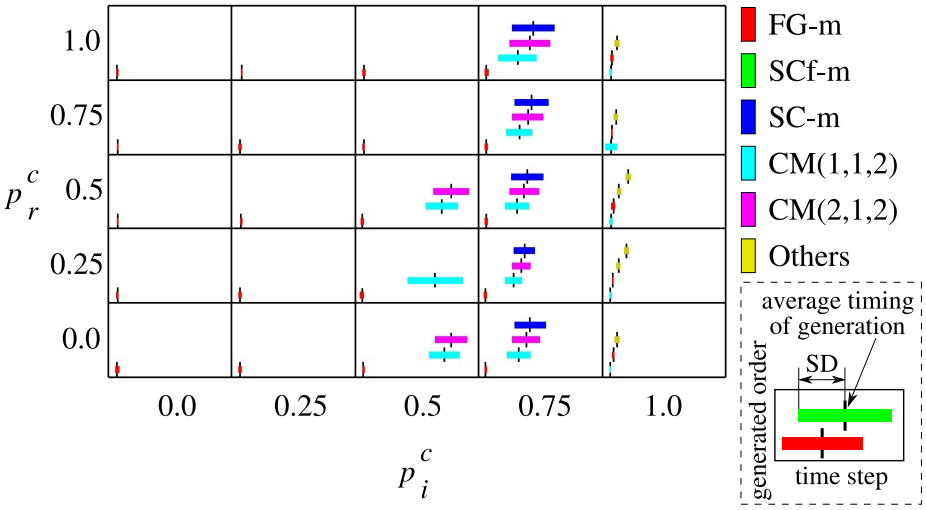
In a naturalistic interaction between a caregiver and an infant, the behavior of the caregiver can be different from the one simulated in the previous section. We examined to what extent the sequence of acquired actions depends on the behavior of the caregiver.

In the simulations, p_r^c and p_i^c were set to either of 0.0, 0.25, 0.5, 0.75, and 1.00 while p_e^c was set to 0.0 or 1.0. If we set $p_e^c = 1.0$, the robot can expect to look at the caregiver's frontal face when it shifts its gaze to the caregiver after gaze following for the caregiver but cannot if $p_e^c = 0.0$. For each parameter setting, we ran ten 100,000-step simulations.

Fig. 5.4 shows the sequence of acquired actions in the case of $p_e^c = 1.0$ (Fig. 5.4(a)) and $p_e^c = 0.0$ (Fig. 5.4(b)). Each section in Figure 5.4 shows the average timing when



(a) a case of $p_e^c = 1.0$.



(b) a case of $p_e^c = 0.0$.

Figure 5.4: The timing of generating CMs under different parameter sets (p_r^c, p_i^c, p_e^c) in face-to-face interactions between a caregiver and a robot.

new CMs were generated. Note that in this analysis, we pick up only CMs that were generated more than five simulations under each parameter set. The horizontal axis in a block indicates time step. The median in a colored rectangle denotes the average and its width represents the standard deviation. A colored rectangle about a CM is stacked in the generated order. We can see that FG-m is first generated under most of parameter sets at almost same time step regardless of the value of p_e^c. A main difference between values of p_e^c is the types of CMs which are generated after FG-m. In the case of $p_e^c = 1.0$, the robot acquired SCf-m and SC-m in the same order as shown in the previous section under most of parameter sets. However the robot could not acquire SC-m if p_r^c was high while p_i^c was low (Figure 5.4(a)).

In the case of $p_e^c = 0.0$, on the other hand, the other CMs were generated as next module of FG-m under some parameter sets (Figure 5.4(b)). CM$(2, 1, 2)$ found in the case with larger p_i^c seems to be another version of shifting the gaze to the caregiver: it enabled the robot to shift the gaze to the caregiver when it was looking at somewhere on the table or the frontal face of the caregiver. CM$(1, 1, 2)$ that was generated before CM$(2, 1, 2)$ constituted a sensori-motor map by which the robot kept looking at the caregiver when it established eye contact with the caregiver. These CMs had contingent connection with FG-m, but they did not have such connection with each other: using output from the CM$(1, 1, 2)$ did not have any positive influence on generation of CM$(2, 1, 2)$, such as promoting increase of $C^1_{2,2}$ although using output from SCf-m promoted increase of $C^1_{\Pi_1,2}$ in the case of $p_e^c = 1.0$ as shown in the previous section.

These results indicate that a caregiver should often shift the gaze to a robot after achieving joint attention with a robot if the caregiver wants it to acquire gaze alternation. We also confirmed the high value of p_e^c promotes generation of SCf-m and SC-m by experiments with setting p_e^c as either of 0.25, 0.5, and 0.75.

5.4 Summary and discussion

In this chapter, we proposed a mechanism to enable a robot to developmentally acquire social actions based on finding and reproducing contingency inherent in face-to-face interaction by the contingency measure based on transfer entropy [23]. We

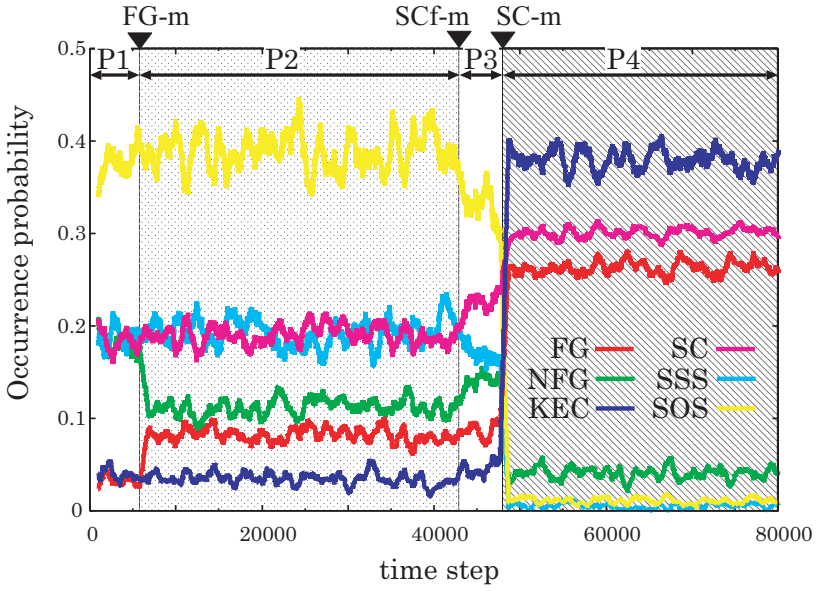


Figure 5.5: Change of the robot's behavior in face-to-face interactions between caregiver and robot.

confirmed that a robot sequentially acquires gaze alternation after acquiring gaze following in computer simulation.

Similarity of developmental order The order of acquiring gaze following and alternation in the experiments is one of remarkable results. In previous studies about acquiring gaze following, gaze alternation was pre-programmed [19] or acquired before acquiring gaze following [20]. However, previous studies in developmental psychology suggest that human infants do not shift their gaze to the caregiver even if they acquire gaze following, but, as they grow, they often shift their gaze to the caregiver [11]. The developmental process of acquiring gaze following and alternation in the experiment is similar to the one of infants. Reproducing contingency inherent in interaction with the caregiver may play an important role in acquiring actions related to joint attention.

Relationship between gaze following and gaze alternation We examined the change of interaction between a caregiver and a robot from the viewpoint of what

actions appear in the interaction as well as of what types of CMs were generated. Figure 5.5 shows an example of change of the frequency of robot actions through interactions with a caregiver with a parameter set $(p_r^c, p_i^c, p_e^c) = (0.5, 0.5, 1.0)$. Here, we separated robot actions into two groups, corresponding to different situations: the first group consists of actions after looking at the caregiver while the second one consists of those after looking at another target. Furthermore, each group was divided into three actions: for the situation of looking at the caregiver following the gaze of a caregiver (FG), not following the gaze of a caregiver (NFG), and keeping eye contact (KEC); while for the situation of looking at other target shifting the gaze to the caregiver (SC), shifting the gaze to the same spot on the table (SS), and shifting the gaze to the other spot on the table (SO). We calculated occurrence rate for each index in interaction during last 1,000 steps.

Interestingly, gaze following for the caregiver and looking at the caregiver after gaze following promoted little change in the robot's behavior (P2 and P3 in Figure 5.5) while looking at the caregiver regardless of gaze following changed the robot's behavior drastically (P4 in Figure 5.5). We can see that the gaze alternation promotes following the caregiver's gaze (red curve in P4 of Figure 5.5) as well as looking at the caregiver (blue and pink curves in P4 of Figure 5.5). This transition might explain conflict of the observation in the developmental process of infant: the observation in laboratory experiments suggests that 6-month-old infants can follow the other's gaze to some extent [10], while caregivers feel that their infants show neither gaze following nor gaze alternation until about ten month of age [15].

Autonomous development An important point of the proposed mechanism is that it enables a robot to acquire social actions sequentially without explicit instructions from a caregiver. This means that a robot can develop them continuously by itself. In the future, adding other action modalities such as pointing or vocalization and sensory modalities to perceive other information about a caregiver such as hand gesture or voice of the caregiver would allow us to examine the relation between other kinds of social actions. Furthermore, we did not consider internal states such as emotion yet. Analyzing the elements with including such states might give us hints to understand how to infer intentions of others.

Modeling of autism Some of previous synthetic approaches [20; 66] have extended their models to autistic models. We can also use the proposed mechanism to understand the developmental process of autistic infants. Some autistic children perform little or no eye contact and tend to avoid looking at human faces. We can realize a model of an autistic infant by using these tendencies as behavior policies in reactive behavior modules. Analyzing contingencies found by a model of a typical infant and the autistic model might provide new understanding of the difference between developmental processes of infants with and without autism. Additionally, we can investigate how the caregiver's behavior influences the learning of several actions related to joint attention for infant models with autism. This might give us new way to improve joint attention skills in autistic infants.

Improvement toward biologically plausible mechanism In human brain, initiating joint attention such as pointing and responding to joint attention such as gaze following seems processed in different areas [9]. Understanding the development of these areas is one of issues to be tackled in neuroscience. We cannot give any suggestion for information processing in human brain during the developmental process of them yet since our mechanism is not based on any findings in neuroscience. We will utilize mechanisms to find contingency in human brain to propose biologically plausible mechanism as a future work.

Chapter 6

Acquisition of gaze following through real-time natural interaction with a human caregiver

6.1 Introduction

In a previous chapter, we proposed that a learning mechanism that enables a robot to acquire several forms of joint attentional behavior and showed the effectiveness of the proposed mechanism using computer simulations. We must, however, determine to what extent the proposed mechanism detects contingency in real-world interaction. Therefore, as the first step to adopt the proposed mechanism to a real robot, we address an issue of how a robot acquires gaze following through real-time interaction with a human caregiver that causes *asynchronous problem*.

The design issues of gaze following are "where and when to shift the gaze," and the existing approaches have been focusing on only "where" issue by assuming the turn taking of gaze change between a human and a robot. Further, they can be classified into two categories: with and without external evaluation. In the former, reinforcement learning [18] or probabilistic algorithms [21] with task evaluation from a supervisor are utilized. In this category, the supervisor always needs to evaluate the robot's behavior. On the other hand, the second category [80; 19] does not need

such evaluation by utilizing the consistency of the relationship between the other agent's gaze direction and the location of a salient object. Especially, Nagai *et al.* introduced a contingency learning mechanism which enabled a robot to acquire gaze following by learning sensory-motor mappings from a human face pattern to own motor command to gaze at an object [19]. However, these approaches are implemented in only computer simulation or not in real-time when applied to real robots. Further, they have not considered "when" issue by synchronizing turn taking of gaze changes between a human and a robot. In order to realize natural interaction between them, real-time interaction without a synchronization assumption should be considered. The issue is how to decide when to shift the gaze to achieve the joint attention with a human.

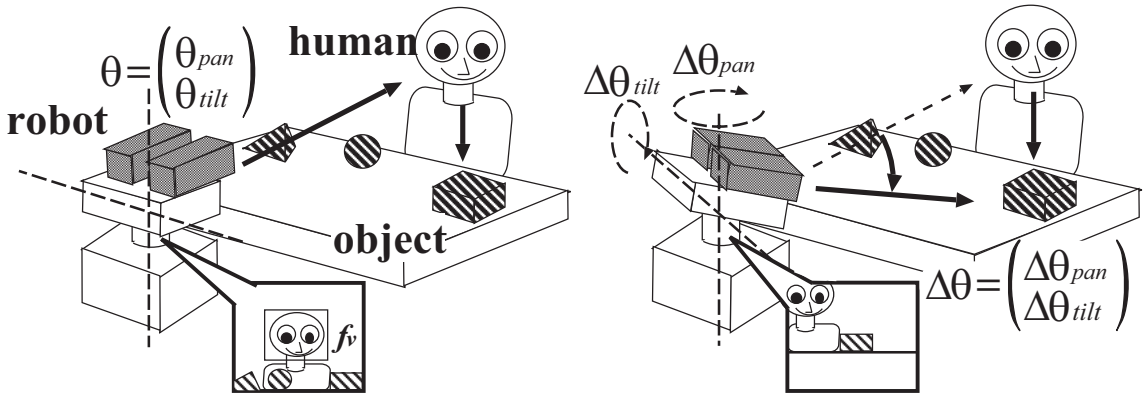
In this chapter, we present a method that solves the issue by introducing an attention selector based on a measure consisting of saliencies of object features and motion information. In order to realize natural interaction that means real-time response without constrained synchronization of gaze shift between a human and a robot, self-organizing map (SOM) for real-time face pattern discrimination [24] and contingency learning for gaze following without external evaluation are utilized. The attention selector controls the robot gaze to switch often from the human face to an object and vice versa, and pairs of a face pattern and a gaze motor command are input to the contingency learning. The motion cues are expected to reduce the number of the incorrect training data pairs due to the asynchronous interaction that affects the convergence of the contingency learning [19].

The rest of this chapter is organized as follows. First, we describe the task of gaze following between a human and a robot, and the problem addressed in this chapter. Next, we give a learning architecture with an attention selector. Then, experimental results on a real robot are given. Finally, we discuss future issues and conclude the chapter.

6.2 gaze following between a human and a robot

The task environment of gaze following between a human and a robot is shown in Figure 6.1. Here, suppose that the robot knows what kind of sensory-motor mapp

it should learn to acquire gaze following. The robot is sitting in front of the human, and there are some objects between them. The robot looks at the human's face pointed to an object and then captures the face pattern $\boldsymbol{f_v}$ from its camera image (see Figure 6.1 (a)). According to the face pattern $\boldsymbol{f_v}$, the robot calculates its head motion $\boldsymbol{\Delta\theta} = (\Delta\theta_p, \Delta\theta_t)$ to turn its head to the object (see Figure 6.1 (b)). Note that a face pattern $\boldsymbol{f_v}$ does not directly indicate an orientation of the face. To achieve gaze following, therefore, the robot needs to learn the sensory-motor mappings from $\boldsymbol{f_v}$ to $\boldsymbol{\Delta\theta}$.



(a) The human looks at an object, and the robot captures a face image pattern, $\boldsymbol{f_v}$.

(b) Based on $\boldsymbol{f_v}$, the robot outputs a motor command $\boldsymbol{\Delta\theta}$ to gaze at the same object the human is looking at (the success of gaze following).

Figure 6.1: gaze following between a robot and a human.

In the previous work [19], this sensory-motor mapping was learned through interactions where the timing of gaze shift between the robot and a human was constrained to ensure consistency of the relation between a human face pattern and positions of the object that the human is looking at: that is, the human needs to tell the robot when to shift its gaze. Since we aim at more natural interactions between a human and a robot, we like to relax such a constraint. If each other's gaze shift is asynchronous, the relationship between a human face pattern and the robot's motor command is not always consistent. This means that it becomes difficult for the robot to learn the sensory-motor mappings because the number of pairs of the incorrect training data

increases. To learn the sensory-motor mapping to perform gaze following, the robot needs to shift its gaze when the consistency of the relation between a human face pattern and positions of objects is ensured.

6.3 The learning architecture utilizing motion cues

Instead of human instructor to tell the robot when to shift its gaze, we utilize motion cues to synchronize the turn taking of gaze change between the human and the robot. The proposed architecture is shown in Figure 6.2, where two key components are 1) an attention selector that decides which face or one of objects to gaze at and when to turn its head utilizing motion information, and 2) an online contingency learning module that enables to acquire gaze following by a spatial contingency within a certain time period [24].

Saliency filters extract different features from the captured camera images. Based on these features (including motion information), an attention selector decides where and when to gaze at. The position of a target (x, y) in the robot's view is sent to the visual feedback module (VFM) that outputs a motor command to gaze at the object. At the same time, an online contingency learning module (LM) outputs another motor command based on similarities between the captured face pattern and pre-categorized face patterns contained in a SOM, and on the robot's posture θ at that time. A gate selects one of these commands and then the robot behaves according to the selected motor command $\Delta\theta$.

6.3.1 Learning process

The robot shifts its gaze to the human's face or a salient object selected by the attention selector (described in the next section). Note that the robot is not programmed to direct its gaze alternately to the human's face and one of objects. Instead, the attention selector decides both which the human face or one of objects to gaze at and when to shift the robot's gaze. It is designed to regard the face as the most salient object because infants are supposed to have innate preference to human faces [32]. Consequently, the robot more often shifts its gaze between the human's face and an

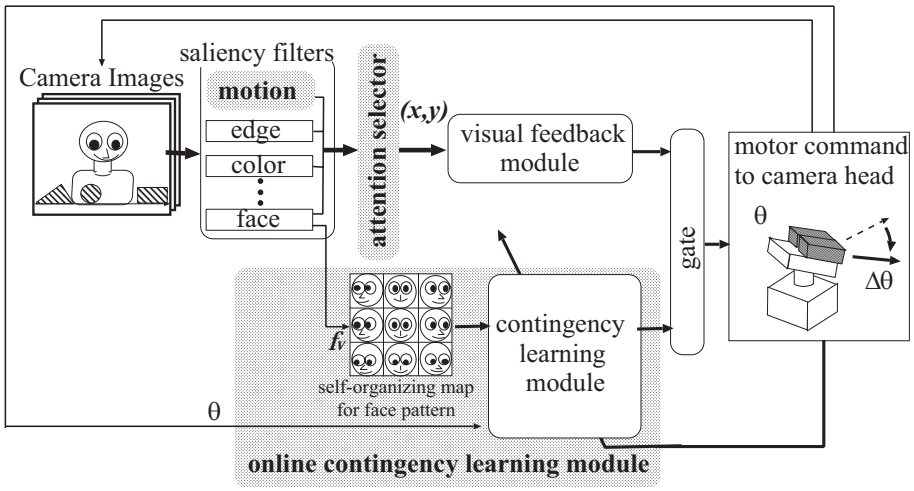


Figure 6.2: An architecture for learning of gaze following through natural interactions based on motion cues.

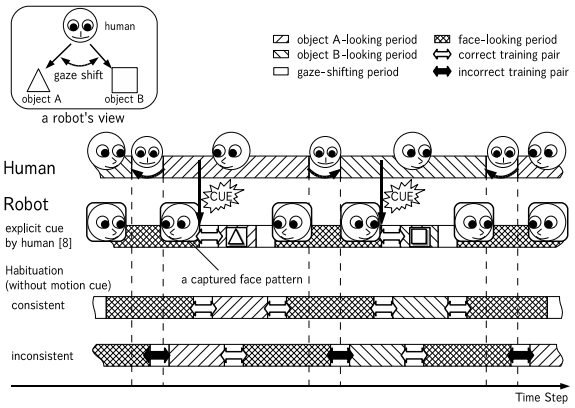
object.

The robot learns the sensory-motor mapping from the face patterns to the motor commands in an almost the same manner as in the previous work [19] but with an attention selector. Now, let the robot gaze at the human's face, and capture the face pattern. Then, it turns its head whenever triggered by the attention selector that utilizes motion cues as one of triggers to shift its gaze. The gate decides whether the robot adopts output from the online contingency learning module (LM) as motor command or not. We use a predetermined sigmoid function as the gate to represent the selecting rate of LM. At the beginning of learning, the gate selects the output from the visual feedback module (VFM) as the robot's motor command and the robot will turn its head to the most salient object that is determined by the attention selector. When it succeeds in gazing at the object around the center of the view, it strengthens the connection between the last face pattern obtained before shifting its gaze and the motor command to gaze at the object regardless of which output the gate selects. Here, this process also occurs in the case of gaze shift from an object to the face to have double chances to obtain the number of the training data pairs and, as a result, it is expected to accelerate learning of gaze following. As learning proceeds, the gate gradually comes to adopt the output from the LM more than one from the VFM.

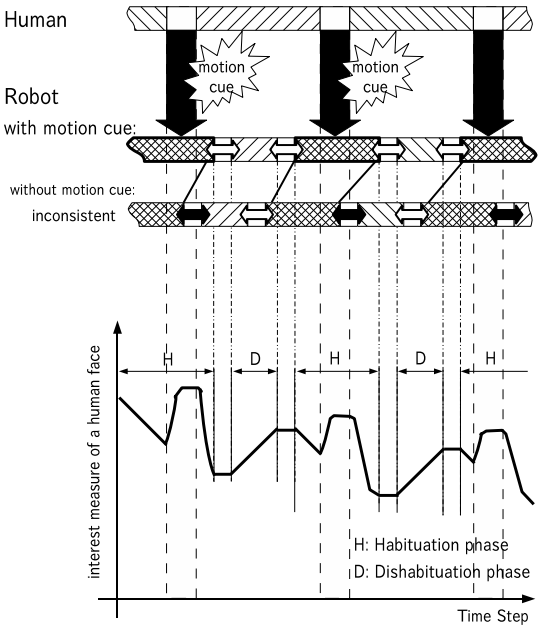
6.3.2 An attention selector

(a) How it works

Although the previous architecture [19] included a mechanism to shift its gaze to the most interesting object, a robot was not able to shift its gaze to another object automatically without any cue from a human after gazing at the object (see Figure 6.3 (a)). In order to realize unconstrained interaction, we introduce an attention selector that is designed based on the phenomenon called habituation in developmental psychology. Habituation can be explained such that human infants lose the interest when they perceive the same stimulus for a while. Therefore, infants change their gaze directions to another stimulus. Some robotics researchers also point out that it is needed for the development of joint attention [80; 81; 65]. We define an interest measure for each object based on image features to model the habituation. The attention selector selects an object according to its selection probability that depends on the interest measure for the object. The higher the probability is, the more often the object is selected to gaze at. As the robot gazes at it, the measure gradually decreases, and then the robot shifts its gaze to another object that has higher interest measure than the current object.



(a) gaze shift triggered by human in Nagai *et al.* [19] and performed by a selector without utilizing motion cues.



(b) gaze shift performed with motion cues and transition of the interest measure of the human face.

Figure 6.3: Effects of motion information on the time periods of the robot's gaze shifts.

Habituation enables a robot to shift its gaze automatically. However, at the same time, it is necessary for the robot to find the periods when the human is gazing at an object and when to shift its gaze to the human or the object to learn gaze following through natural interaction with a human. "Natural interaction" means real-time response with unconstrained synchronization of gaze shift between the human and the robot. This is important especially in the case that the human moves the object that he/she is looking at to a different location: in such a case, since the robot needs the time to change its gaze, the robot may miss the timing to capture the correct pair of the human face pattern and the position of object. For example, Figure 6.3 (a) shows a simple example to indicate the difference between the gaze shifts triggered by a human and by an attention selector based on habituation but without motion cues.

In Figure 6.3, it is assumed that a human shifts the gaze alternately to the objects A and B at a constant frequency and a robot looks at the object A, a face, and the object B in order. Note that the robot captures a human face pattern both before its gaze shift from the human face to an object and after from an object to the face [1]. With the attention selector, there are two cases; the case where only correct pairs of the face pattern and the motor command are input to the learning system and the case where incorrect pairs are included: if the robot shifts its gaze during gaze shift by a human, it cannot learn the correct relation between the last face pattern obtained before shifting its gaze and the motor command output to gaze at an object (see solid both–side arrows).

To solve this problem, we construct the interest measure including not only object–specific image features, such as color and edge, but also motion information such as a human head turn or motions of objects manipulated by the human. In developmental psychology, there are some observations that an infant shifts its gaze utilizing an adult's head turn or the moving hand as well as motion of objects [25; 36] as one of cues of gaze shift. Therefore, this implementation is appropriate as a human infant model. Shifting the gaze based on motion cues enables a robot to change the timing of the gaze shift depending on the timing when the human shifts the gaze

[1] It captures a face pattern only before it shifts its gaze from the face to an object in previous work [19].

and picks up an object. Then, we designed the parameters of attention selector in a such a way that motion cue causes the rapid increase of the interest measure of a moving object or a turning face. If a robot gazes at the moving object or face, the robots gaze at it longer. If the robot does not, the robot shifts its gaze to it immediately. Figure 6.3 (b) indicates a simple example in the case where a robot gaze at the turning face. The top shows changes of robot's gaze shift based on motion cues. The bottom shows transition of the interest measure of the human face, where H and D indicate habituation and dishabituation phases, respectively. Note that interest measures between phases do not change because they are not calculated when the robot rotates its head. Motion cues about a human head turn increase the interest measure of the face, and the robot keeps gazing at the face until the human stops turning the head. As a result, an attention selector with motion information can provide a robot with more chances to obtain the correct training data pairs in the inconsistent case of Figure 6.3 (a) than one without motion information, and acceleration of learning gaze following is expected.

(b) The mechanisms of the attention selector

The robot can extract a human's face image by detecting a face-like area, and extract objects by detecting object-specific features such as color and edge. These image features, including the face-like one, are candidates for the robot to gaze at.

Let n be the number of candidates for objects to be looked at in the robot's camera image. The interest measure $I_i(t)$ of each candidate is defined as

$$I_i(t) \quad = \quad M_i(t)S_i(t) \quad (i = 1, 2, \cdots, n, n+1), \tag{6.1}$$

where t is the sampling time, and the $(n+1)$-th candidate shows the interest measure of the human's face. $I_i(t)$ consists of the motion saliency, $M_i(t)(> 0)$, and the object-specific saliency, $S_i(t)(> 0)$. $M_i(t)$ denotes a value that is influenced by how long the i-th candidate moves until the sampling time t and is defined as

$$M_i(t) \quad = \quad g(m_i(t)), \tag{6.2}$$

where $m_i(t)$ represents the degree of motion and is defined as follows:

$$m_i(t) = \begin{cases} m_i(t-1)+1 & (|\boldsymbol{f}_i| > \epsilon_i) \\ max\{m_i(t-1)-1, 0\} & (|\boldsymbol{f}_i| \leq \epsilon_i) \end{cases}, \tag{6.3}$$

where the flow vector $\boldsymbol{f}_i$ for the i-th candidate is calculated by optical flows, and ϵ_i is a small positive constant. Motion detection is prohibited when the robot rotates its head to avoid the confusion of motion detection due to its own motion or independent object motions.

In equation (6.2), the function g is a kind of threshold function. Here we use the following function:

$$g(x) \;=\; 1 + \frac{a}{1+\exp\{(d-x)/T\}}, \tag{6.4}$$

where a, d, T are positive real numbers. The parameter a decides influence of motion information on the interest measure. The larger a is, the higher the probability of selection for the i-th candidate is when it moves. The parameter d is set to absorb noise about the flow vector and T decides the sensitivity to motion information. We set each parameter in terms that the function enables a robot to detect both human face motion and objects'.

The motion saliency, $M_i(t)$, changes the gaze duration of a robot, such as "object-A-looking period", "object-B-looking period" and "face-looking period" in Figure 6.3 significantly. If a human turns the head when the robot is gazing at the face, an attention selector with motion cues detects the timing of a human head turn and the motion saliency about the human face $M_{n+1}(t)$ increases. As a result, the robot keeps looking at the face until the head turn stopping because the interest measure about the human face $I_{n+1}(t)$ is increasing. This increase realizes the motion synchronization of shifting the gaze between the robot and the human to obtain the correct training data.

$S_i(t)$ shows the object-specific saliency of the i-th candidate. We set an initial value of $S_i(t)$ as

$$S_i(0) \;=\; C_i \quad (C_i > 0), \tag{6.5}$$

where C_i is a weighting constant to decide the basic bias for the robot to select the i-th candidate, that is, a preference to the candidate. We initialize the larger value of the

human face than other objects' so that the robot can simulate the innate preference of infants to human faces [32]. $S_i(t)$ is defined as follows:

$$S_i(t+1) = \begin{cases} \alpha_i S_i(t) & \text{if the i-th candidate is attended} \\ max\{C_i, \beta_i S_i(t)\} & \text{else} \end{cases}, \tag{6.6}$$

where α_i $(0 < \alpha_i < 1)$ is a decay factor while β_i (> 1) is a growth factor. Equation (6.6) means the object-specific saliency $S_i(t)$ gradually decreases during the robot continues to gaze at the i-th candidate and vice versa. The decay and growth factors for a candidate influence habituation and dishabituation phases, respectively as shown in the bottom of Figure 6.3 (b). It is expected that the robot shifts its gaze more frequently than the human because it can have the more opportunities to learn training data pairs in the situation where the consistency of the relationship between the other agent's gaze direction and the location of a salient object is ensured. The robot also needs to experience shifting the gaze alternately to the human's face and one of objects as much as possible to learn gaze following. Therefore, the interest measure of the human face should be designed to decrease and recover faster than the measures of other objects.

The robot calculates the interest measure $I_i(t)$ for each candidate. According to the interest measures, the selection probability $\Pr(i,t)$ for the i-th candidate is calculated as follows:

$$\Pr(i,t) = \frac{I_i(t)}{\sum_{j=1}^{n+1} I_j(t)}. \tag{6.7}$$

Note that the human's face and objects are not distinguished in the target selection process though they are different in learning process. Therefore, the robot sometimes shifts the gaze from one object to another or keeps gazing at the same target.

6.3.3 An online contingency learning module

An online contingency learning module strengthens the connection between a face pattern and own motor command to turn its head to an object. The point of this learning process is that the human does not provide the robot with any evaluation whether or not the connection is appropriate to acquire gaze following. In addition,

the robot cannot explicitly find which object the human looks at. That is, through the learning process, the contingency learning module strengthens not only relevant connections but also irrelevant ones. Nagai *et al.* [19], however, shows if positions of objects change randomly and the human gazes at objects, the relevant connections to acquire gaze following are more strengthened than irrelevant ones because there exists a contingency between a face pattern and the position of the object that the human looking at. As a result, the robot can acquire gaze following based on this contingency. We leave the details to Nagai *et al.* [19].

Instead of the high-dimensional face image matching [19] that consume a large amount of computation, we utilize a SOM of face patterns [24]. In advance, we make a robot learn the SOM to categorize face patterns in which each neuron represents a vector of a gray scale face image. As inputs of learning of gaze following, we utilize the activations of each neuron calculated based on the similarity with a face image that the robot is gazing at. Figure 6.4 shows a network, where two-layered perceptron with an SOM input layer is learned through backpropagation by utilizing the robot's motor command as reference signal. The compression of input dimension by the SOM enables the robot to discriminate face pattern in real time.

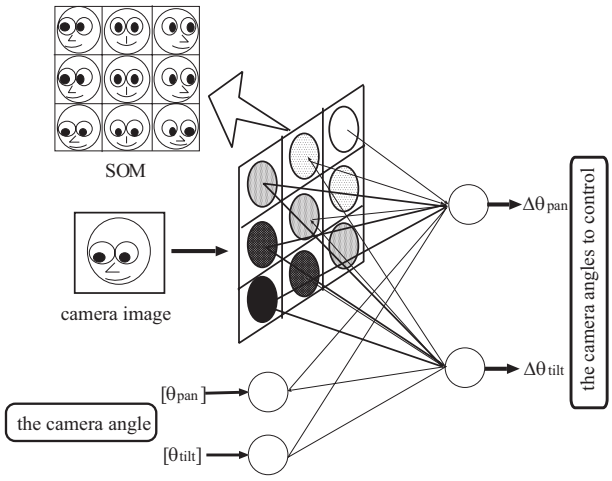


Figure 6.4: Online contingency learning module: a robot learns the relation between the activations of individual neurons in the SOM calculated based on the similarity with the captured face pattern and its motor command.

6.4 Experiments

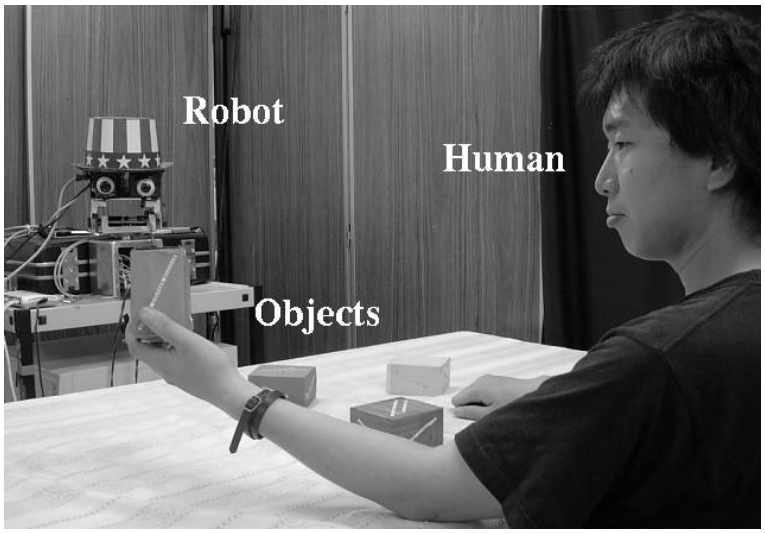


Figure 6.5: The experimental setting: the robot and the human are seated face-to-face and between them there are four objects with different colors.

6.4.1 Environmental setup

The experimental setup is shown in Figure 6.5. The robot and the human are seated face-to-face. Throughout the experiment, the distance between a human and a robot is constant. Four objects with different colors are placed on the table between them. The robot head has two degrees of freedoms (DOFs): the pan and tilt. A CCD camera (Firefly produced by Point Grey) on the head provides 320×240 color video images at 30 frames per second. Note that the horizontal and vertical angles of view of the camera are about 61.9 and 48.5 degrees, respectively, and these angles are wide enough for the robot to capture both the human face and objects on the table. The template matching method is used as face detector and a 32×32 pixel face-like region is extracted. Also, the color areas are extracted as object regions, and an optical flow by the block matching method is detected.

The robot learned the SOM to categorize face patterns within three minutes before

it learns joint attention. Figure 6.6 shows a learned SOM used in the later gaze following learning. The learned SOM consists of 9×9 clusters, each of which is constituted by a 32×32 pixel gray scale image based on the face-like region extracted by the face detector.

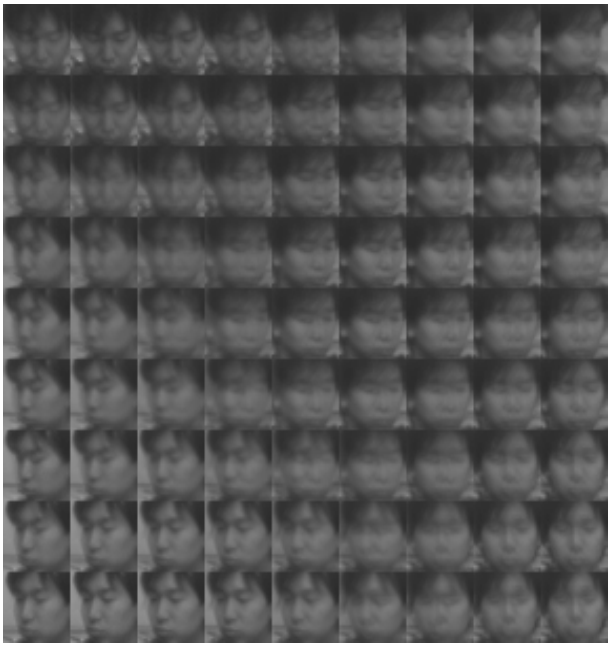


Figure 6.6: A learned SOM of the face patterns.

In the following experiments, we assume the robot can always observe a human face and some objects in the field of view. Table 6.1 shows parameters used in the attention selector, and parameters in the threshold function $g(x)$ (eq. (6.4)) were set as $(a, d, T) = (4.5, 20, 1.4)$. The robot took about one second to shift its gaze from one target to another.

In addition, we used a sigmoid function as a gate. The robot decides whether it adopts the output from the online contingency learning module as a motor command according to the probability Pr_g:

$$Pr_g(l) = \frac{1}{1.0 + \exp\{(p - l)/q\}}, \tag{6.8}$$

where l is the number of learning iteration. As the learning proceeds, Pr_g becomes

higher. As a result, the robot gradually comes to adopt the output from the learning module. Each of parameters in the gate function decides learning time. Before experiments, therefore, we performed preliminary experiments to determine the parameters of the gate in an environment where the robot could learn most easily, that is, the timing of gaze shift between the human and the robot is synchronized completely. Based on the result, parameters of Pr_g were set as $(p, q) = (150, 22.5)$ (see Figure 6.7). This represents selecting rate of learning module's output reaches to 50% at the 150th learning step.

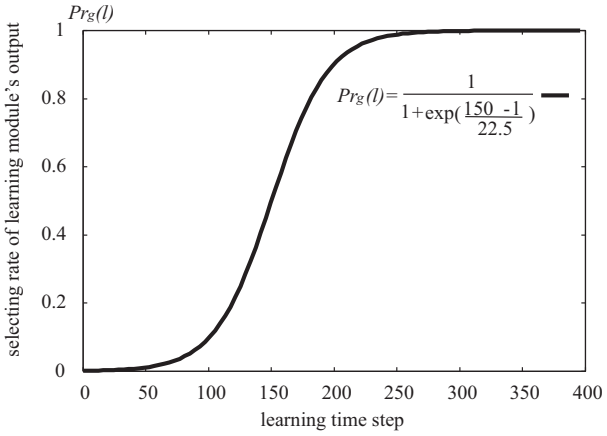


Figure 6.7: The gating function used in the experiments.

Table 6.1: parameters of attention selector.

Candidate	C_i	α_i	β_i
the human's face	1500	$\exp(-2.0 \times 10^{-2})$	$\exp(1.2 \times 10^{-2})$
the object: A (red), B (yellow), C (blue), D (green)	800	$\exp(-1.0 \times 10^{-2})$	$\exp(2.0 \times 10^{-3})$

6.4.2 Human behavior

The task for the human is the object-transfer task: a person (here, a male) randomly selects one object and directs his gaze to it. Next, he picks it up and observes it for about two seconds, and then, puts it somewhere on the table, gazes at it for about two seconds, and selects another object. Note that the object manipulated by him is arranged in different positions of the table as evenly as possible and a moving object is only what he is manipulating. The robot also shifts its gaze to one of the objects and the person's face according to the decision of its attention selector.

6.4.3 Learning gaze following

We investigated whether the robot could acquire gaze following through human–robot natural interaction. To validate an effect of motion cues, we compared performances between the architectures with and without motion cues five times. Note that the architecture without motion cues utilizes only object-specific features to select a target. In each session, we counted whether the robot was able to perform joint attention with the human or not when it directed its gaze from his face to an object. Each session lasted approximately 26 minutes. The average number of the robot's gaze shift was 302.0 times with motion cues and 279.6 times without them. The standard deviations were 6.87 and 8.36, respectively. Also, the average numbers of success of gaze following were 199.6 and 124.4, respectively and the standard deviations were 6.25 and 19.26, respectively.

Figure 6.8 shows the averages and standard deviations in five sessions of moving averages of the success rate of gaze following in terms of with/without motion cues. Each moving average at a given time t minutes, $mov\_ave(t)$, in one experiment was calculated as follows:

$$mov\_ave(t) = \frac{\text{number of gaze following from t-1 to t+1}}{\text{number of robot's gaze shift from t-1 to t+1}}. \tag{6.9}$$

In Figure 6.8, '$\times$' and '+' indicate the results with and without motion cues, respectively. The vertical bar at each point represents the standard deviation of five sessions. Note that the success rate at the beginning of learning includes the success of gaze following by visual feedback. While, the success rate at the end of learning

indicates the performance by the online contingency learning module. We can see the gaze shift with motion cue significantly improves the performance over without motion cue, and the success rate of gaze following by the proposed architecture reaches 80% after about 20 minutes. Most of failures happened when the robot gazes at a distractor very close to the target. Here, distractors mean the other objects that the subject does not gaze at.

Although the subject did not exactly behave in the same manner, we observed the same tendencies in five sessions in spite that a robot experienced different timing and frequency of human gaze shift. The results with other subjects also showed the same tendencies. Therefore, the proposed architecture may have the validity in experimental environment.

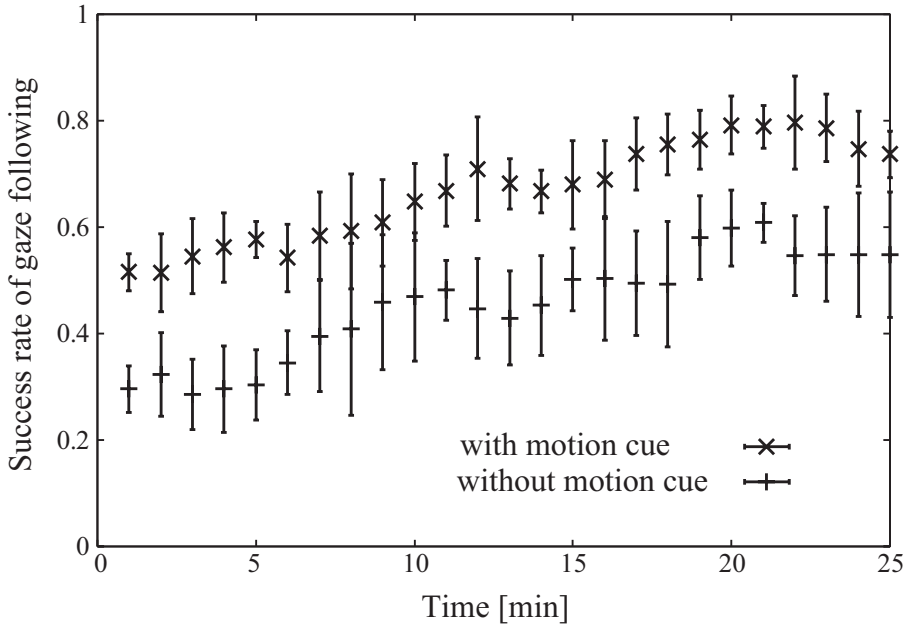


Figure 6.8: The time courses of success rate of gaze following through interaction with a human.

6.5 Summary and discussion

In this chapter, we concentrated on acquisition of gaze following through real-time interaction as the first step to adopt the mechanism proposed in the chapter 5 to a real robot. In such interaction, the robot confronts *asynchronous problem*: it has to know when to shift its gaze to acquire gaze following. Here, we presented a method that solves this problem by introducing an attention selector based on a measure consisting of saliencies of object features and motion information. The experimental result showed the gaze shift utilizing motion cues enables a robot to acquire gaze following efficiently.

Biological plausibility In our approach, we utilized the motion information expecting to accelerate learning of gaze following, and we obtained successful results. In developmental psychology, it is suggested as one of precursors of gaze following that 3– and 4– month-olds, who do not have an ability of gaze following with adults, often shift their gaze to adults' moving hand and/or an object in their hand [36]. Therefore, it is plausible that shifting the gaze based on motion information increases the chances to obtain the consistent training data pairs and helps infants to acquire joint attention.

Improvement of performance As mentioned in the previous section, most failures are caused by distractors near by the target in the image. If they were distant from the target in 3–D space, these failures might have been avoided by using the depth cues from binocular vision system. In addition, if the robot knows the human attention strategy model through interactions with him/her, the robot might be able to find the target correctly. Hoffman *et al.* propose a model that enables a robot to learn instructor–specific saliency models by performing gaze following with a human but they need the evaluation for robot's behavior [21]. Without such evaluation, we should build a learning model that can acquire both gaze following and an ability to infer other's preference.

Learning of synchronization In our experiments, the designer specified the parameters of habituation such as how long the robot gazes at an object. These parameters should be estimated for the robot to be synchronized with human behaviors (head turn and object transfer) through real interactions. Carlson and his colleagues propose a simulation model that can synchronize the gaze shift by a reinforcement learning method [80; 81], and we may apply their method to estimate the parameters for the synchronization.

Utilizing temporal structure The attention selector directly utilized the motion cues in the object–transfer task. In the behavior such as the object manipulation, it is supposed that coordination of eye and hand movements has a temporal structure [82]. Therefore, such a structure might be useful for more accurate synchronization due to the capability of prediction of motion sequences. Furthermore, if the pace of each motion can be estimated through interactions, more adaptive synchronization might be possible depending on situations. Actually, the caregivers may change the paces of their motions to adapt themselves with children's behaviors [83; 84; 85].

Analysis of human–robot interaction In Figure 6.3 (b), the motion cue is used in one way from the human to the robot, but actually the human caregiver is also affected by the robot behavior. In Figure 6.8, the performance without motion cues appears to have slightly improved by accident due to this effect. We need to observe human–robot interaction with and without this effect and utilize the result to build a robot that can acquire shared gaze following through more natural interaction with a human.

Chapter 7

Conclusion and future work

In this dissertation, we dealt with building a robot that acquires various forms of joint attentional behavior. The following issues were addressed in previous chapters.

- The issue of chapter 4 was how the robot can find the contingency in the interaction. An information theoretic measure was proposed to find a contingent structure in face-to-face interaction between a caregiver and a robot. The measure consists of transfer entropy that is an information theoretic measure representing the flow of information between stochastic variables. We showed that it enables a robot to find a contingent relationship between a variable of face pattern and a variable of shifting its gaze utilized used for learning gaze following in computer simulations of face-to-face interaction.

- In chapter 5, we proposed a learning mechanism that developmentally acquires various forms of joint attention related actions based on the proposed measure in chapter 4. The mechanism constructed sensory-motor mapping from a state of a sensory variable to a most informative action of a action variable in the found contingent pair of sensory and action variables. The robot showed behavior based on the already acquired sensory-motor mapping to reproduce the found contingency. That further led novel contingency from the interaction with a caregiver. The results of computer simulations indicated that a robot acquires a series of actions related to joint attention in an order that almost matches with an infant's development of joint attention. In addition, we indicated that

looking back at a caregiver may be important behavior to utilize gaze following.

- As the first step to apply the proposed mechanism to a real robot, we considered that the real-time interaction without an assumption of turn taking of gaze change between a caregiver and a robot in chapter 6. We addressed the issue of how to decide when to shift the gaze for learning gaze following with a human. We introduced an attention selector based on a measure consisting of saliencies of object features and motion information. The motion cues provided a robot with when a human caregiver shift the gaze. The experimental result shows the gaze shift utilizing motion cues enables a robot to synchronize its own motion with human motion and to learn gaze following efficiently in about 20 minutes.

7.1 Future work

In each chapter, we discussed some future works. We did not, however, describe issues to be addressed to apply the proposed mechanism to a real robot. Our mechanism is influenced by interaction between a caregiver and a robot. To validate our mechanism, we need to implement our mechanism into a real robot. Therefore, we discuss future works to apply the proposed mechanism to a real robot in this section.

Rapid mechanism for finding contingency It is an important topic to find a contingency quickly because a human caregiver has difficulty in keeping natural interaction with a robot for a long time. Our mechanism needs too many interactions at least 45,000 steps for acquiring gaze alternation as shown in the chapter 5. In order to counter this problem, we should improve the proposed measure by using more effective measure such as effective transfer entropy [86].

Resolution of random variables The resolution of the random variables may influence the estimation of SAC concerning the variables though we set it in advance. An infant seems to be faced with situations in which the resolutions of multimodal sensation or various kinds of action are different because these components develop in parallel and according to a different time schedule. The resolutions of random variables would improve incrementally along with the infant's development. Therefore,

we should address the issue of how a robot can improve such resolutions and maintain the development of social skill based on the detection of the contingencies.

Reproducing contingency in continuous space A contingency reproduction module introduced in chapter 5 was composed of a sensory-motor map based on discrete data set of sensory information and motor command. In order to apply our mechanism in real-time interaction, we need to deal with continuous data set of them. One solution is that after the contingency detector finds a contingency for an event, it learns a sensory-motor map using the continuous data set. The contingency reproduction module can know which combination of states in the variables is informative by contingent saliencies of the event. Therefore, we expect that a contingency reproduction module can learn a sensory-motor map based on continuous data set by regarding the most informative combination as a reward.

Bibliography

[1] T. Fong, I. Nourbakhsh, and K. Dautenhahn. *A survey of socially interactive robots.* Robotics and Autonomous Systems **42**(3-4), 143 (2003).

[2] C. Breazeal. *Social interactions in HRI: the robot view.* IEEE Transactions on Systems, Man and Cybernetics Part C **34**(2), 181 (2004).

[3] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. *Development and evaluation of interactive humanoid robots.* Proceedings of the IEEE (Special issue on Human Interactive Robot for Psychological Enrichment) **92**(11), 1839 (2004).

[4] L. Adamson. *Communication development during infancy* (Westview Press, 1996).

[5] M. Tomasello. *The cultural origins of human cognition* (Harvard University Press, 1999).

[6] G. E. Butterworth. *Origins of mind in perception and action.* In M. C. and P. J. Dunham, eds., *Joint attention: It's origins and role in development*, chap. 2, pp. 29–40 (Lawrence Erlbaum Associates, 1995).

[7] C. Moore and P. J. Dunham, eds. *Joint attention: Its origins and role in development* (Lawrence Erlbaum Associates, 1995).

[8] J. Seibert, A. Hogan, and P. Mundy. *Assessing interactional competencies: The early social-communication scales.* Infant Mental Health Journal **3**(4), 244 (1982).

[9] P. Mundy, J. Card, and N. Fox. *Eeg correlates of the development of infant joint attention skill.* Developmental Psychobiology **36**, 325 (2000).

[10] G. E. Butterworth and N. L. M. Jarrett. *What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy.* British Journal of Developmental Psychology **9**, 55 (1991).

[11] M. Tomasello. *Joint attention as social cognition.* In C. Moore and P. J. Dunham, eds., *Joint attention: It's origins and role in development*, chap. 6, pp. 103–130 (Lawrence Erlbaum Associates, 1995).

[12] S. Baron-Cohen. *Mindblindness* (MIT Press, 1995).

[13] C. Moore. *Theories of mind in infancy.* British Journal of Developmental Psychology **14**(19-40) (1996).

[14] V. Corkum and C. Moore. *Development of joint visual attention in infants.* In C. Moore and P. J. Dunham, eds., *Joint attention: It's origins and role in development*, chap. 4, pp. 61–84 (Lawrence Erlbaum Associates, 1995).

[15] H. Ohgami. *The developmental origins of early joint attention behaviors.* Kyushu Univ. Psycho. Res. **3**, 29 (2002). In Japanese.

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

[17] R. Pfeifer and C. Scheier. *Understanding Intelligence* (The MIT Press, 1999).

[18] G. Matsuda and T. Omori. *Learning of joint visual attention by reinforcement learning.* In *International Conference on Cognitive Modeling*, pp. 157–162 (2001).

[19] Y. Nagai, K. Hosoda, A. Morita, and M. Asada. *constructive model for the development of joint attention.* Connection Science **15**(4), 211 (2003).

[20] J. Triesch, C. Teuscher, G. O. Deák, and E. Carlson. *Gaze following: why (not) learn it?* Developmental Science **9**(2), 125 (2006).

[21] M. Hoffman, D. Grimes, A. Shon, and R. Rao. *A probabilistic model of gaze imitation and shared attention.* Neural Networks **19**(3), 299 (2006).

[22] S. Watson, J. *Smiling, cooing, and "the game".* Merrill-Palmer Quarterly **18**(4), 323 (1972).

[23] T. Schreiber. *Measuring information transfer.* Phys. Rev. Lett. **85**(2), 461 (2000).

[24] A. Morita, Y. Yoshikawa, K. Hosoda, and M. Asada. *Joint attention with strangers based on generalization through the joint attention with caregivers.* In *the IEEE/RSJ International Conf. on Intelligent Robots and Systems*, pp. 3744–3749 (2004).

[25] C. Moore, M. Angelopoulos, and P. Bennett. *The role of movement in the development of joint visual attention.* Infant Behavior and Development **20**, 83 (1997).

[26] M. Scaife and J. Bruner. *The capacity for joint visual attention in the infant.* Nature **253**, 265 (1975).

[27] M. Morales, P. Mundy, and J. Rojas. *Following the direction of gaze and language development in 6-month-olds.* Infant behavior and development **21**(2), 373 (1998).

[28] D. A. Baldwin and E. M. Markman. *Establishing word-object relations: A first step.* Child Development **60**, 381 (1989).

[29] D. A. Baldwin. *Infants' contribution to the achievement of joint reference.* Child Development **62**, 373 (1991).

[30] R. Brooks and A. Meltzoff. *The development of gaze following and its relation to language.* Developmental Science **8**(6), 535 (2005).

[31] F. Sai and I. W. R. Bushnell. *The perception of faces in different poses by 1-month-olds.* British journal of developmental psychology **6**, 35 (1988).

[32] J. G. Bremner. *Infancy: 2nd Edition* (Oxford: Blackwell, 1994).

[33] T. Farroni, G. Csibra, F. Simion, and M. H. Johnson. *Eye contact detection in humans from birth.* Proceedings of the National Academy of Sciences **99**(14), 9602 (2002).

[34] L. Cohen and G. Amsel. *Precursors to infants' perception of the causality of a simple event.* Infant Behavior and Development **21**(4), 713 (1998).

[35] S. Desrochers, P. Morissette, and M. Ricard. *Two perspectives on pointing in infancy.* In M. C. and P. J. Dunham, eds., *Joint attention: It's origins and role in development*, chap. 5, pp. 85–102 (Lawrence Erlbaum Associates, 1995).

[36] S. Amano, E. Kezuka, and A. Yamamoto. *Infant shifting attention from an adult's face to an adult's hand: a precursor of joint attention.* Infant Behavior and Development **27**, 64 (2004).

[37] M. Kuroki. *The effect of positive emotion on infants' gaze shift.* Infant Behavior and Development **30**, 606 (2007). Issue 4.

[38] P. Rochat. *The infant's world*, chap. 4, pp. 127–166 (Harvard University Press, 2001).

[39] A. L. Woodward. *Infants' developing understanding of the link between looker and object.* Developmental Science **6**(3), 297 (2003).

[40] R. Flom, G. O. Deák, C. G. Phill, and A. D. Pick. *Nine-month-olds' shared visual attention as a function of gesture and object location.* Infant Behavior and Development **27**, 181 (2004).

[41] R. Bakeman and L. Adamson. *Coordinating attention to people and objects in mother–infant and peer–infant interaction.* Child Development **55**, 1278 (1984).

[42] M. Carpenter, K. Nagell, and M. Tomasello. *Social cognition, joint attention, and communicative competence from 9 to 15 months of age.* Monographs of the society for research in child development **63**(4) (1998).

[43] G. Gergely. *What should a robot learn from an infant? mechanisms of action interpretation and observational learning in infancy.* Connection Science **15**(4), 191 (2003).

[44] J. S. Watson. *Social Perception in Infants*, chap. Contingency perception in early social development, pp. 157–176 (Ablex Pub. Corp., 1985).

[45] T. Striano, A. Henning, and D. Stahl. *Sensitivity to social contingencies between 1 and 3 months of age.* Developmental Science **8**(6), 509 (2005).

[46] P. Rochat. *The infant's world* (Harvard University Press, 2001).

[47] J. Nadel, I. Carchon, C. Kervella, D. Marcelli, and D. Réserbat-Plantey. *Expectancies for social contingency in 2-month-olds.* Developmental Science **2**(2), 164 (1999).

[48] P. J. Dunham and F. Dunham. *Optimal social structures and adaptive infant development.* In C. Moore and P. J. Dunham, eds., *Joint attention: It's origins and role in development*, chap. 8, pp. 159–189 (Lawrence Erlbaum Associates, 1995).

[49] L. Adamson and R. Bakeman. *Mothers' communicative acts: Changes during infancy.* Infant Behavior and Development **7**, 467 (1984).

[50] T. Kishimoto, Y. Shizawa, J. Yasuda, T. Hinobayashi, and T. Minami. *Do pointing gestures by infants provoke comments from adults?* Infant Behavior and Development **30**(4), 562 (2007).

[51] P. Yoder and L. Munson. *The social correlates of co-ordinated attention to adult and objects in motor-infant interaction.* First Language **15**, 219 (1995).

[52] N. J. Sasson. *The development of face processing in autism.* Journal of Autism and Developmental Disorders **36**(3), 381 (2006).

[53] G. Dawson, J. Munson, A. Estes, J. Osterling, J. McPartland, K. Toth, L. Carver, and R. Abbott. *Neurocognitive function and joint attention ability in young children with autism spectrum disorder versus developmental delay.* Child Development **73**(2), 345 (2002).

[54] P. Mundy, M. Sigman, and C. Kasari. *A longitudinal study of joint attention and language development in autistic children.* Journal of Autism and Developmental Disorders **20**(1), 115 (1990).

[55] M. Siller and M. Sigman. *The behaviors of parents of children with autism predict the subsequent development of their children's communication.* Journal of Autism and Developmental Disorders **32**(2), 77 (2002).

[56] O. I. Lovaas. *Behavioral treatment and normal educational and intellectual functioning in young autistic children.* Journal of Consulting and Clinical Psychology **55**(1), 3 (1987).

[57] M. Imai, T. Ono, and H. Ishiguro. *Physical relation and expression: Joint attention for human–robot interaction.* In *Proceedings of 10th IEEE International Workshop on Robot and Human Communication* (2001).

[58] H. Kozima and H. Yano. *A robot that learns to communicate with human caregivers.* In *The First International Workshop on Epigenetic Robotics* (2001).

[59] A. Ito and K. Terada. *Producing intentionality in eye-contact robot.* In *Proceedings of 11th International Conference on Human-Computer Interaction*, pp. 22–27 (2005).

[60] C. Breazeal and B. Scassellati. *A context-dependent attention system for a social robot.* In *In Proceedings of the Sixteenth International Joint Conference on Artificial Intellignece*, pp. 1146–1151 (1999).

[61] B. Scassellati. *Theory of mind for a humanoid robot.* Autonomous Robots **12**(1), 13 (2002).

[62] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. *Developmental robotics: a survey.* Connection Science **15**(4), 151 (2003).

[63] F. Kaplan and V. V. Hafner. *The challenges of joint attention.* Interaction Studies **7**(2), 135 (2006).

[64] Y. Nagai, M. Asada, and K. Hosoda. *Learning for joint attention helped by functional development.* Advanced Robotics **20**(10), 1165 (2006).

[65] I. Fasel, G. O. Deák, J. Triesch, and J. Movellan. *Combining embodied models and empirical research for understanding the development of shared attention.* In *Proceedings of the 2nd International Conference on Development and Learning*, pp. 21–27 (2002).

[66] C. Teuscher and J. Triesch. *To each his own: The caregiver's role in a computational model of gaze following.* Neurocomputing **70**, 2166 (2007).

[67] B. Lau and J. Triesch. *Learning gaze following in space: a computational model.* In *International Conference on Development and Learning* (2004).

[68] K. Hosoda, H. Sumioka, and M. Asada. *Acquisition of human-robot joint attention through real-time natural interaction.* In *the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2867–2872 (2004).

[69] Y. Nagai. *Joint attention development in infant-like robot based on head movement imitation.* In *Proceedings of the Third International Symposium on Imitation in Animals and Artifacts*, pp. 87–96 (2005).

[70] A. Meltzoff and M. Moore. *Explaining facial imitation: A theoretical model.* Early Development and Parenting **6**, 179 (1997).

[71] T. Konno and T. Hashimoto. *Developmental construction of intentional agency in communicative eye gaze.* In *Proceedings of the Fifth International Conference on Developmental and Learning* (2006). CD-ROM.

[72] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. *Artificial intelligence: Autonomous mental development by robots and animals.* Science **291**(5504), 599 (2001).

[73] A. Barto, S. Singh, and N. Chentanez. *Intrinsically motivated learning of hierarchical collections of skills.* In *International Conference on Development and Learning* (2004).

[74] P.-Y. Oudeyer, F. Kaplan, and V. Hafner. *Intrinsic motivation systems for autonomous mental development.* IEEE Transactions on Evolutionary Computation **11**(2), 265 (2007).

[75] M. Lungarella, K. Ishiguro, Y. Kuniyoshi, and N. Otsu. *Methods for quantifying the causal structure of bivariate time series.* International Journal of Bifurcation Chaos **17**(3), 903 (2007).

[76] O. Sporns, J. Karnowski, and M. Lungarella. *Mapping causal relations in sensorimotor networks.* In *Proc. of the 5th International Workshop on Epigenetic Robotics* (2006).

[77] M. Lungarella and O. Sporns. *Mapping information flow in sensorimotor networks.* PLoS Computational Biology **2**(10), 1301 (2006).

[78] M. Lungarella, T. Pegors, D. Bulwinkle, and O. Sporns. *Methods for quantifying the informational structure of sensory and motor data.* Neuroinformatics **3**(3), 243 (2005).

[79] H. Sumioka, Y. Yoshikawa, and M. Asada. *Causality detected by transfer entropy leads acquisition of joint attention.* In *6th IEEE International Conference on Development and Learning* (2007). CD-ROM, Poster 93.

[80] E. Carlson and J. Triesch. *A computational model of the emergence of gaze following.* In *Proceedings of the 8th Neural Computation and Psychology Workshop, Progress in Neural Processing. World Scientific* (2003).

[81] C. Teuscher and J. Triesch. *To care or not to care: Analyzing the caregiver in a computational gaze following framework.* In *the 3rd International Conference on Development and Learning* (2004).

[82] J. Pelz, M. Hayhoe, and R. Loeber. *The coordination of eye, head, and hand movements in a natural task.* Experimental Brain Research **139**, 266 (2001).

[83] K. J. Rohlfing, J. Fritsch, B. Wrede, and T. Jungmann. *How can multimodal cues from child–directed interaction reduce learning complexity in robots?* Advanced Robotics **20**(10), 1183 (2006).

[84] R. J. Brand, D. A. Baldwin, and L. A. Ashburn. *Evidence for 'motionese' : modifications in mothers' infant-directed action.* Developmental Science **5**(1), 72 (2002).

[85] J. S. Herberg, M. M. Saylor, and D. T. Levin. *The perceived intentionality of an audience influences action demonstrations.* In *Proceedings of the 5th IEEE International Conference on Development and Learning* (2006).

[86] R. Marschinski and H. Kantz. *Analysing the information flow between financial time series. an improved estimator for transfer entropy.* The European Physical Journal B **30**(2), 275 (2002).

Published Papers by the Author

Articles in Journals

1. Hidenobu Sumioka, Koh Hosoda, Yuichiro Yoshikawa, and Minoru Asada. "Acquisition of joint attention through natural interaction utilizing motion cues", *Advanced Robotics*, Vol.21, No.9, pp.983–999, 2007.

2. Hidenobu Sumioka, Yuichiro Yoshikawa, and Minoru Asada. "Learning of Joint Attention from Detecting Causality Based on Transfer Entropy", *Journal of Robotics and Mechatronics*, Vol.20, No.3, pp.378–385, 2008.

3. Hidenobu Sumioka, Yuichiro Yoshikawa, and Minoru Asada. "Reproducing Interaction Causality Toward Open-ended Development of Social Actions: Case Study on Joint Attention", *Advanced Robotics*, (*submitted*)

Papers in Proceedings of International Conferences

1. Hidenobu Sumioka, Koh Hosoda, Yuichiro Yoshikawa, and Minoru Asada, "Motion-triggered human-robot synchronization for autonomous acquisition of joint attention", In *Proceedings of the 2005 4th IEEE International Conference on Development and Learning*, 2005.

2. Hidenobu Sumioka, Yuichiro Yoshikawa, and Minoru Asada, "Causality Detected by Transfer Entropy Leads Acquisition of Joint Attention", In *Proc. of the 6th IEEE International Conf. on Development and Learning*, 2007.

3. Yuichiro Yoshikawa, Shunsuke Yamamoto, Hidenobu Sumioka, Hiroshi Ishiguro, and Minoru Asada, "Spiral Response-cascade Hypothesis–Intrapersonal

Responding–cascade in Gaze Interaction-", In *Proceedings of the 3rd ACM/IEEE International Conference Human-Robot Interaction*, Amsterdam, March, 2008.

4. Hidenobu Sumioka, Yuichiro Yoshikawa, and Minoru Asada, Development of Joint Attention Related Actions Based on Reproducing Interaction Causality, In *Proc. of the 7th IEEE International Conf. on Development and Learning*, 2007 (*to appear*).

Papers in Proceedings of Japanese Conferences and Miscellanies

1. 住岡英信, 細田耕, 吉川雄一郎, 浅田稔. "コミュニケーション相手の動作情報を利用した共同注意の自律的獲得". ロボティクス・メカトロニクス講演会 '05 予稿集, Vol.CD-ROM, 2A1-N-042, 2005.

2. 住岡英信, 細田耕, 吉川雄一郎, 浅田稔, "動き情報を利用した共同注意の自律的獲得", 第５回日本赤ちゃん学会学術集会, P23, 2005.

3. 住岡英信, 吉川雄一郎, 浅田稔. "移動エントロピーによる因果関係の発見とこれに基づく共同注意の獲得". ロボティクス・メカトロニクス講演会 '07 予稿集, Vol.CD-ROM, 1A2-L08, 2007.

4. 住岡英信, 吉川雄一郎, 浅田稔, "母子間相互作用における因果関係の発見が導く社会的行動の発達～ロボットによる視線追従の獲得を例に～", 第７回日本赤ちゃん学会学術集会, pp. 41(P03), 2007.

5. 竹内佑治, 吉川雄一郎, 住岡英信, 浅田稔, "移動経験が導く共同注意発達の構成的モデル", 第７回日本赤ちゃん学会学術集会, pp. 43(P05), 2007.

6. Hidenobu Sumioka, Yuichiro Yoshikawa, Minoru Asada. "Can Causality Detected by Transfer Entropy Lead Acquisition of Gaze Following?". 第１回浅田共創知能システムシンポジウム, P10, 2007.

7. 住岡英信, 吉川雄一郎, 浅田稔. "母子間相互作用における因果関係の発見に基づく共同注意行動の逐次的獲得モデル". 日本赤ちゃん学会第８回学術集会プログラム抄録集, pp.71, 2008.

8. 中野 吏, 吉川 雄一郎, 住岡 英信, 浅田 稔. ”相互排他性に基づくマルチーモダル共同注意～視線とラベルによる共同注意モジュールの相互促進的学習”. 日本赤ちゃん学会　第８回学術集会, pp.64, 2008.

9. 住岡英信, 吉川雄一郎, 浅田稔. ”因果関係の再現に基づく社会的行動の逐次的獲得モデルー視線追従から社会的参照へー”. ロボティクス・メカトロニクス講演会 ’08 予稿集, Vol.DVD-ROM, 2A1-E21, 2008.

10. 山本俊介, 吉川雄一郎, 住岡英信, 石黒浩, 浅田稔. ”自らの接近行動がパーソナルスペースの認知におよぼす効果”. ロボティクス・メカトロニクス講演会 ’08 予稿集, Vol.DVD-ROM, 2P2-I04, 2008.

11. 竹内佑治，住岡英信，吉川雄一郎，浅田稔，”養育者の関わりの変容を考慮した移動を伴う共同注発達モデル”，第 25 回日本ロボット学会学術講演会，3B16，2007

12. 山本俊介，吉川雄一郎，住岡英信，石黒浩，浅田稔, ”応答的視線による観察が他者認知に及ぼす効果”，情報処理学会関西支部　支部大会　講演論文集，75–78，2007