Development of Joint Attention Related Actions Based on Reproducing Interaction Contingency

Hidenobu SUMIOKA[1,2], Yuichiro YOSHIKAWA[1], and Minoru ASADA[1,2]
[1]JST ERATO Asada Synergistic Intelligence Project
[2]Graduate School of Eng., Osaka Univ.
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
Email: {sumioka, yoshikawa, asada}@jeap.org

***Abstract*—Understanding the developmental process of joint attention related actions, such as gaze following and alternation, is one of essential issues for the emergence of communication. Previous synthetic studies have proposed learning methods for gaze following without any explicit instructions as the first step to understand the development of these actions. However, a robot was given *a priori* knowledge about which pair of sensory information and action should be associated. This paper addresses the development of social actions without such knowledge with a learning mechanism that iteratively acquires social actions by finding and reproducing the contingency inherent in the interaction with a caregiver. The measurement of contingency based on transfer entropy is used to find appropriate pairs of variables for acquiring social actions from possible candidates. The reproduction of found contingency promotes a change of contingent structure in the subsequent actions of a caregiver and a robot. In computer simulations of human-robot interaction, we examine what kinds of actions related to joint attention can be acquired in which order by controlling the behavior of caregiver agents. The result shows that a robot acquires joint attention related actions in an order that resembles an infant's development of joint attention.**

***Index Terms*—joint attention, transfer entropy, contingent chain, sequential acquisition of social behavior**

I. Introduction

Joint attention, especially visual joint attention, is defined as looking at the same object that someone else is looking at. It is the most important for attention sharing among agents because where a person is looking at often reveals where the person's attention is being directed. Therefore, acquiring actions related to joint attention such as gaze following, pointing, gaze alternation, and social referencing is a central topic in developmental psychology [1]. Infants seem to acquire various kinds of behavior gradually in the development of joint attention; they begin to show gaze alternation, i.e., successive looking between a caregiver and an object, social referencing, and pointing after learning gaze following [2]. However, it remains unclear why most infants acquire several forms of joint attention related behavior in such an order.

In robotics, joint attention studies have been done not only from the viewpoint of building communicative robots [3] but also from synthetic approaches to modeling human developmental processes. Previous synthetic studies have addressed how infants can acquire gaze following without explicit instructions about where to look [4], [5]. A contingent structure has been shown in a sequence of gazing actions of an infant robot and its caregiver. The structure enables it to successfully acquire gaze following based on their statistical association of these actions. However, the robot had *a priori* knowledge about which pair of perception and action variables should be associated thought it learned a sensory-motor mapping to reproduce the contingent relation. Communicative robots usually have many candidates for motor and perceptual variables to be associated to acquire such social actions because they are supposed to have multimodal sensory-motor experiences that reflect the contingent interaction with a human. This indicates that it is not trivial for a robot to select such a pair of sensory-motor variables to model contingency involved in the interaction.

We regard such contingency as how predictive certain pair of sensory-motor variables is of next sensory state. In order to determine to what extent there is high contingent relationship between variables [7], we use the transfer entropy measure [6]. 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, transfer entropy was confirmed to be useful for the robot to find an appropriate combination of variables that enables learning of joint attention [8]. Since infants seem to use the contingency inherent in interactions with their caregivers to acquire social skills [9], such a measure of contingency is expected to be useful for acquiring not only gaze following but also other kinds of joint attention behavior. We hypothesize that showing behavior based on the acquired sensory-motor mapping, that is 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. We expect to model developmental process of joint attention by finding contingency and its reproduction.

In this paper, we propose a mechanism that iteratively acquires social actions by extending the proposed measure in a previous work [8] in such a way that the mechanism does not only find a combination of contingent variables but also constructs a sensory-motor mapping to reproduce behavior based on the found contingency. In addition, it adds a new variable expressing whether the constructed sensory-motor mapping is used to promote further finding of novel contingency depending on already found contingency. In computer simulations, we examine which order of developing actions related to joint attention is generated depending on what type of caregiver. It selects an appropriate sensory-motor mapping used to decide the next action from the previously-acquired mappings. In computer simulations, we examine which order of developing actions related to joint attention is generated depending on the type of caregiver.

In order to show the validity of the proposed method, the computer simulation of face-to-face interaction is conducted, and future issues are discussed.

II. Contingency inherent in a social interaction

We focus on face-to-face interaction between a caregiver and a robot who, in each time step, take turns observing the opponent and its surround each other as follows. First, the caregiver acts, and then the robot observes the scene to obtain N_S types of sensory information $\boldsymbol{S} = \{S_1, S_2, \cdots, S_{N_S}\}$ called sensory variables. After that, it chooses N_A types of actions from the $\boldsymbol{A} = \{A_1, A_2, \cdots, A_{N_A}\}$ called action variables, and then observes N_R types of the resultant sensory information $\boldsymbol{R} = \{R_1, R_2, \cdots, R_{N_R}\}$ called the resultant sensory variables. We call a triplet (S_i,A_j,R_k) an event. Here, the contingency inherent in the interaction appears as a statistical bias in a certain event. The robot's task is to find a contingent event from possible ones and to learn a sensory-motor mapping from a sensory variable to a motor one in the selected event to reproduce the contingency.

Each of social actions consists of a sequence of sub-actions, such as social referencing where one might follow the other's gaze to find an object and then look back to the face of the other to ask why the other looks at it. This implies that a new social action will be acquired by reproducing chains of contingent relationship. Therefore, the found social actions can be used not only for deriving successive and/or more sophisticated social actions from caregivers but also for seeking for the chains of contingent relationship. Finding the chains is promoted by treating the actions as higher level sensory and motor variables.

III. A mechanism to successively develop social behaviors

Fig. 1 shows a mechanism that enables a robot to acquire social actions based on finding contingency inherent in interaction. 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 the beginning of learning. They are generated by the contingency detector through interaction with a caregiver.

RMs and CMs output not only motor commands to be executed but also their reliability values for the current state. A reliability value reflects the reproducibility of an expected resultant sensory experience by a motor command for the current state. The module selector determines a motor command

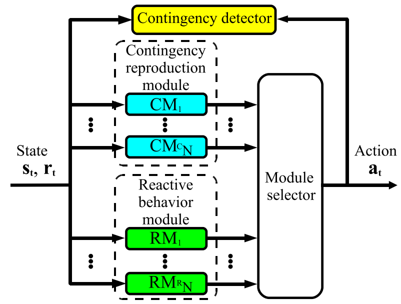


Fig. 1. Proposed mechanism to successively develop social actions

based on the reliability. A pair of the state and the selected commands are used by the contingency detector to create a new CM as well as to locate a contingent event.

A. Contingency detector

A contingency detector has two main roles: finding a contingent event and generating a new CM based on it. We propose an information theoretic measure of contingency based on transfer entropy [6] to quantify the contingency in events experienced through interactions with a caregiver. Transfer entropy is the information measure that represents the flow of information between stochastic variables. The contingency detector evaluates contingency in interaction based on 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

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

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 S_i and A_j on R_k, we introduce sensory-motor contingency (SMC) $C^j_{i,k}$, which is defined and expanded as follows:

$$\begin{aligned} C^j_{i,k} &= T_{(S_i,A_j)\to R_k} - (T_{S_i\to R_k} + T_{A_j\to R_k}) \\ &= \sum_{\substack{s^t_i \in S_i, \\ r^t_k \in R_k}} p(r^t_k, s^t_i) \sum_{\substack{r^{t+1}_k \in R_k, \\ a^t_j \in A_j}} e(r^{t+1}_k, a^t_j | r^t_k, s^t_i), \quad (2) \end{aligned}$$

where $e(r^{t+1}_k, a^t_j | r^t_k, s^t_i)$ in Eq. (2) is called a contingent saliency under a pair of observables (r^t_k, s^t_i).

$$\begin{aligned} e(r^{t+1}_k, a^t_j | r^t_k, s^t_i) = p(r^{t+1}_k, a^t_j | r^t_k, s^t_i) \log \frac{p(r^{t+1}_k | r^t_k, s^t_i, a^t_j)}{p(r^{t+1}_k | r^t_k, s^t_i)} \\ -p(r^{t+1}_k, a^t_j | r^t_k) \log \frac{p(r^{t+1}_k | r^t_k, a^t_j)}{p(r^{t+1}_k | r^t_k)}. \quad (3) \end{aligned}$$

$e(r_k^{t+1}, a_j^t | r_k^t, s_i^t)$ represents a statistical bias on state transition from r_k^t to r_k^{t+1} originating from a pair of s_i^t and 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. (3) becomes larger. However, there is a possibility that the difference derives from a dependency only on a pair a_j^t and r_j^t. Therefore, the second term in Eq. (3), which represents influence of only a_j^t on 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 SMCs for all triplets, the detector evaluates whether to generate a new CM for an event with the highest SMC value. Here, a new CM for a SMC is generated if its SMC keeps the highest value during T^C time steps and exceeds a threshold θ between the last consecutive steps. Hereafter, a CM that is constructed for an event (S_i, A_j, R_k) is denoted as $\mathrm{CM}(i, j, k)$.

When the contingency detector creates the i-th new CM, the set of events is extended by adding new variables, A_{Π_i} and S_{Π_i}. A_{Π_i} represents whether output from a CM is used as a current motor command to perform its current action while S_{Π_i} represents whether output from the 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 SMCs $C_{i,k}^j$ 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 action variables, respectively.

B. Contingency reproduction module

A CM is composed of a sensory-motor map from a sensory variable to an action variable of the found event. 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:

$$(r_k^*, a_j^*) = \operatorname*{argmax}_{r_k', a_j'} e(r_k', a_j' | r_k, s_i), \tag{4}$$

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

$$Z(r_k, s_i) = \frac{e_{max}(r_k, s_i) - \mu^{s_i, r_k}}{\sigma^{s_i, r_k}}, \tag{5}$$

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} are the average of the contingent saliencies under observables (r_k, s_i) and the standard deviation, respectively.

However, not every pair of observables are 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: we check the averages (μ and μ_σ) of all averages and standard deviations under pairs of observables. If the average μ^{r_k, s_i} and the standard deviation σ^{r_k, s_i} under observables (r_k, s_i) exceed μ and μ_σ, the reliability under the observables is calculated.

C. Reactive behavior module

A RM outputs an action based on the behavior policies preprogrammed 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 a reliability too. Here, the value of α is designed to allow a robot to select outputs from CMs.

D. 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 for this purpose. The module selector determines a motor command for an action from outputs of the CMs and RMs based on their reliabilities.

Let ${}^C N$ and ${}^R N$ denote the numbers of CMs and RMs that output motor commands for the i-th action, respectively. Here, we express ${}^C N^i$ CMs and ${}^R N^i$ RMs as $\mathrm{M}_j^i \in \{\mathrm{RM}_1^i, \mathrm{RM}_2^i, \cdots, \mathrm{RM}_{{}^R N^i}^i, \mathrm{CM}_1^i, \mathrm{CM}_2^i, \cdots, \mathrm{CM}_{{}^C N^i}^i\}$. Probability $\Pr(\mathrm{M}_j^i)$ of choosing the output from M_j^i at the t-th step is given by softmax selection such as:

$$\Pr(\mathrm{M}_j^i) = \frac{\exp(Z^j(r^t, s^t, t')/\tau_i)}{\sum_{l \in {}^R N^i + {}^C N^i} \exp(Z^l(r^t, s^t, t')/\tau_i)}, \tag{6}$$

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, it is designed to select different action within a few steps. The parameter τ_i is the temperature parameter. If τ_i is set as a constant value, the increase of ${}^C N^i$ decreases the $\Pr(\Gamma_j^i)$. Therefore, we decrease τ_i as ${}^C N^i$ increases: $\tau_i = 1/({}^C N^i + 1)$

E. 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. Through iteration of interaction between a caregiver and the robot the contingency detector selects a contingent event and generates a new CM that constructs a sensory-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, we expect a new CM for an event involving S_Π or A_Π can be generated because we assume each of social actions consists of a sequence of sub-actions. As a result, the robot acquires actions that are related to each other.

IV. Computer simulation of developing joint attention related behavior

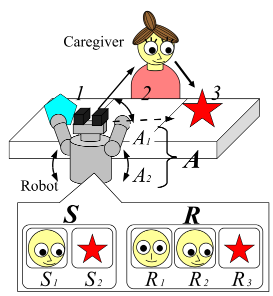


Fig. 2. Experimental setting for acquisition of joint attention behaviors

The performance of the proposed model was tested in computer simulations where an infant model (hereafter robot) interacted with a caregiver model in face-to-face situations. While the basic environmental setting follows the one in the previous work [8], the robot selected actions based on the proposed mechanism and caregiver actions were simulated more faithfully.

A. Experimental setting

1) Environment and Infant model: Fig. 2 shows an overview of the setting in the computer simulation. The robot sits across from the caregiver at a fixed distance. 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).

TABLE I
Initial variables in robot

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

The variables in Table I were set as initial variables in the robot model. Here, we assume that the robot can always observe the states of these variables correctly. The direction of the caregiver's gaze is denoted by S_1 of which member indicates the gaze to a particular location of a table (f_1, f_2, f_3) or the robot (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 of which member indicates that the robot is looking at an object (o) or something else (o_ϕ). We prepare three types of variables as resultant ones: caregiver's full 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 of which member 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 of which member indicates the four different hand gestures. Here, parameters about the proposed mechanism were set as $(T^C, \theta, \alpha, \beta) = (30, 5.0 \times 10^{-5}, 1.0, 0.5)$.

2) The behavior rules of the caregiver model: In the previous study [8], we modeled the caregiver's behavior so that she not only randomly looks at the robot or at one of the objects but also shows responsive and inductive behavior. Here, we adopt a similar model for the caregiver, except that the current model shows responsive behavior 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 (Responding to joint attention process; in short, RJA process); 2) shifting gaze to draw the robot's attention (Initiating joint attention process; IJA process); 3) randomly selecting a target to gaze at (neutral process) excluding behavior identical to RJA and IJA processes. If she is looking at the robot's face, RJA process is selected with probability p^c_r and the neutral process is selected otherwise. If she is looking at an object, IJA process is selected with probability p^c_i and the neutral process is selected otherwise. In addition, the caregiver shifts the gaze to the robot's face with probability p^c_e if she and the robot successfully look at the same object.

B. Sequential acquisition of joint attention behavior

We ran 100,000 steps simulations ten times where the parameters were set as $(p^c_r, p^c_i, p^c_e) = (0.5, 0.5, 1.0)$. The average number of CMs found by the contingency detector was 3.8. In 80 % 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$(\Pi_1, 1, 2)$. 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$(\Pi_1, 1, 2)$; hereafter **SC-m**). Moreover, they were often generated earlier than other CMs for different events.

Fig. 3 shows examples of time courses of SMCs for several events which have ever been one of the two highest SMC values through 80,000 steps in a simulation. Any events related to hand gesture have no high value because hand gesture does not involve any contingency in this interaction. The vertical

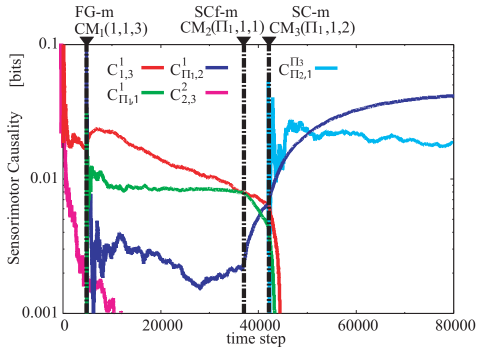


Fig. 3. Time courses of the sensory-motor contingency of events in a simulation of face-to-face interactions between a caregiver and a robot.

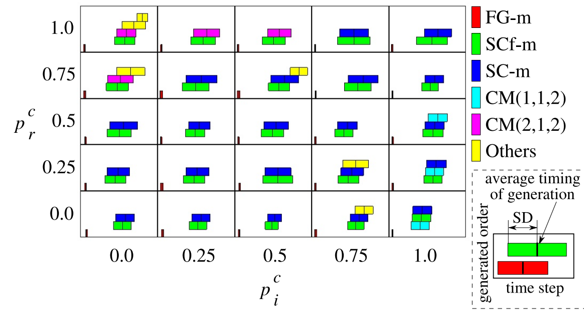
axis indicates the logarithmic value of the SMCs. We also show the timing of generating new CMs as inverted black triangle on the top of the graph. After interactions were iterated, $C^1_{1,3}$ first became the highest among all SMCs (red curve in Fig. 3). As a result, the FG-m was generated at the 4825-th step. 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 sensory-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. (3), smaller. This 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 36794-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 contingent structure in interaction again and promoted increase of $C^1_{\Pi_1,2}$ (blue curve in Fig. 3). This caused the generation of SC-m for the event (S_{Π_1}, A_1, R_2) at the 42054-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 a robot acquired not only gaze following but also gaze alternation through the loop of finding and reproducing a chain of contingency in interaction that change by using output from existing CMs.

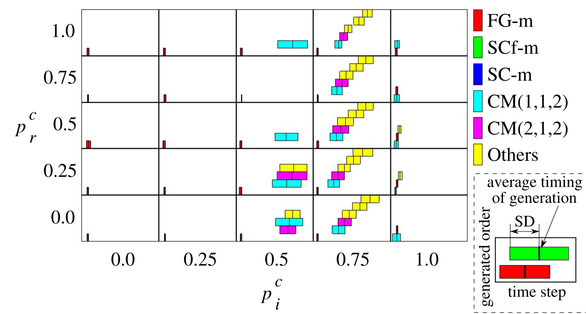
C. Influence of caregiver's behavior

In more realistic situation 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^c_r and p^c_i were set to either of 0.0, 0.25, 0.5, 0.75, and 1.00 while p^c_e was set to 0.0 or 1.0. If we set $p^c_e = 1.0$, the robot can expect to look at the caregiver's full face when it shifts its gaze to the caregiver after gaze following for the caregiver but cannot if $p^c_e = 0.0$. For each parameter setting, we ran 100,000-step simulations ten times.



(a) a case of $p^c_e = 1.0$.



(b) a case of $p^c_e = 0.0$.

Fig. 4. The timing of generating CMs under different parameter sets (p^c_r, p^c_i, p^c_e) in face-to-face interactions between a caregiver and a robot.

Fig. 4 shows the sequence of acquired actions in the case of $p^c_e = 1.0$ (Fig. 4(a)) and $p^c_e = 0.0$ (Fig. 4(b)). Each section in Fig. 4 shows the average timing when 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 of each section of the figures 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^c_e. A main difference between values of p^c_e is the types of CMs which are generated after FG-m. In the case of $p^c_e = 1.0$, 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^c_r was high while p^c_i was low (Fig. 4(a)).

In the case of $p^c_e = 0.0$, on the other hand, the other CMs were generated as next module of FG-m under some parameter sets (Fig. 4(b)). $\mathrm{CM}(2, 1, 2)$ found in the case with larger p^c_i seems to be a limited version of shifting the gaze to the caregiver: it enabled the robot to shift the gaze to the caregiver when it was looking at a spot on the table or the full face of the caregiver. $\mathrm{CM}(1, 1, 2)$ that was generated before $\mathrm{CM}(2, 1, 2)$ constructed a sensory-motor map by which the robot kept looking at the caregiver when it established eye contact with the caregiver. These CMs had causal 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 SC-m in the case of $p^c_e = 1.0$ 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^c_e promotes generation of SCf-m and SC-m by experiments with setting p^c_e as either of 0.25, 0.5, and 0.75.

V. Discussion and conclusion

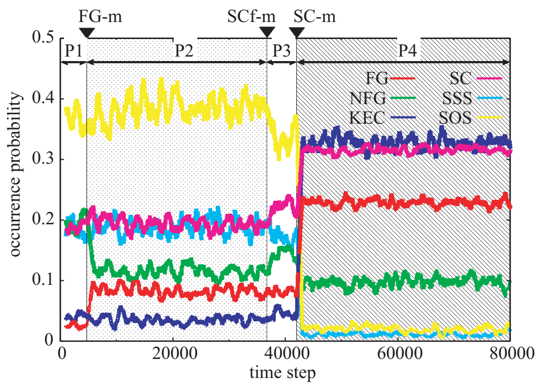


Fig. 5. Change of the robot's behavior of face-to-face interactions between caregiver and robot in an example of a simulation.

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 proposed measure in the previous work [8]. We confirmed that a robot sequentially acquires gaze alternation after acquiring gaze following in computer simulation.

An interesting point of this result is that a robot acquires joint attention related actions in an order that resembles an infant's development of joint attention. In previous studies about acquiring gaze following, gaze alternation was preprogrammed [4] or acquired before acquiring gaze following [5]. 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 [2]. The developmental process of acquiring gaze following and alternation in the experiment seems similar to the one of infants. Reproducing contingency inherent in interaction with the caregiver may play an important role in acquiring joint attention related actions.

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. Fig. 5 shows an example of change of the frequency of robot actions through interactions with a caregiver with a parameter set $(p^c_r, p^c_i, p^c_e) = (0.5, 0.5, 1.0)$. Here, we parted 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 Fig. 5) while looking at the caregiver regardless of gaze following changed the robot's behavior drastically (P4 in Fig. 5). We can see that the gaze alternation promotes following the caregiver's gaze (red curve in P4 of Fig. 5) as well as looking at the caregiver (blue and pink curves in P4 of Fig. 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 seem to feel that their infants show neither gaze following nor gaze alternation until about ten month of age [11].

The proposed mechanism enables a robot to acquire social actions sequentially without explicit instructions from a caregiver. 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.

References

[1] Moore, C. and Dunham, P. (Eds.), *Joint attention: It's origins and role in development*, Lawrence Erlbaum Associates (1995).

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

[3] Imai, M., Ono, T. and Ishiguro, H., Physical Relation and Expression: Joint Attention for Human-Robot Interaction, In *Proc. of 10th IEEE International Workshop on Robot and Human Communication*, (2001).

[4] Nagai, Y., Hosoda, K., Morita, A., and Asada, M., A constructive model for the development of joint attention, *Connection Science*, **15**(4), pp.211–229, (2003).

[5] Triesch, J., Teuscher, C., Deák, G., and Carlson, E., Gaze following: why (not) learn it?, *Developmental Science*, **9**(2), pp.125–157, (2003).

[6] Schreiber, T., Measuring information transfer, *Phys. Rev. Lett.*, **85**(2), pp.461–464, (2000).

[7] Sporns, O., Karnowski, J., and Lungarella, M., Mapping Causal Relations in Sensorimotor Networks, In *Proc. of the 5th International Workshop on Epigenetic Robotics*, (2006).

[8] Sumioka, H., Yoshikawa, Y., and Asada, M., Causality Detected by Transfer Entropy Leads Acquisition of Joint Attention, In *Proc. of the 6th IEEE International Conf. on Development and Learning*, (2007).

[9] Watson, J., Smiling, cooing, and "the game", *Merrill-Palmer Quarterly*, **18**(4), pp.323–339, (1972).

[10] Butterworth, G. and Jarrett, N., What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy, *British Journal of Developmental Psychology*, **9**, pp. 55–72, (1991).

[11] Ohgami, H., The developmental origins of early joint attention behaviors, *Kyushu Univ. Psycho. Res.*, **3**, pp.29–39, (2002) , (In Japanese)