# How does an infant acquire the ability of joint attention?: A Constructive Approach

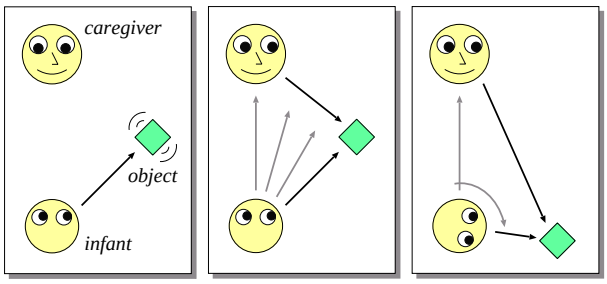
**Yukie Nagai*** **Koh Hosoda*†** **Minoru Asada*†**
*Dept. of Adaptive Machine Systems,
†HANDAI Frontier Research Center,
Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
yukie@er.ams.eng.osaka-u.ac.jp, {hosoda, asada}@ams.eng.osaka-u.ac.jp

## Abstract

This study argues how a human infant acquires the ability of joint attention through interactions with its caregiver from the viewpoint of a constructive approach. This paper presents a constructive model by which a robot acquires a sensorimotor coordination for joint attention based on visual attention and learning with self-evaluation. Since visual attention does not always correspond to joint attention, the robot may have incorrect learning situations for joint attention as well as correct ones. However, the robot is expected to statistically lose the data of the incorrect ones as outliers through the learning, and consequently acquires the appropriate sensorimotor coordination for joint attention even if the environment is not controlled nor the caregiver provides any task evaluation. The experimental results suggest that the proposed model could explain the developmental mechanism of the infant's joint attention because the learning process of the robot's joint attention can be regarded as equivalent to the developmental process of the infant's one.

## 1. Introduction

A human infant acquires various and complicated cognitive functions through interactions with its environment during the first few years. However, the cognitive developmental process of the infant is not completely revealed. A number of researchers (Bremner, 1994, Elman et al., 1996, Johnson, 1997) in cognitive science and neuroscience have attempted to understand the infant's development. Their behavioral approaches have explained the phenomena of the infant's development, however its mechanisms are not clear. In contrast, constructive approaches have potential to reveal the cognitive developmental mechanisms of the infant. It is suggested in robotics that the building of a human-like intelligent robot based on the insight of the infant could lead us to the understanding of the mechanisms of the infant's development (Brooks et al., 1998, Asada et al., 2001).




(a) 6th month (b) 12th month (c) 18th month

Figure 1: Development of infant's joint attention

Joint attention with a caregiver is one of the abilities that help the infant to develop its social cognitive functions (Scaife and Bruner, 1975, Moore and Dunham, 1995). It is defined as a process that the infant attends to an object which the caregiver attends to. Owing to the ability of joint attention, the infant learns other kinds of social functions, e.g. language communication, mind reading (Baron-Cohen, 1995), and so on. On the basis of the insight, robotics researchers have attempted to build the mechanisms of joint attention for their robots (Breazeal and Scassellati, 2000, Scassellati, 2002, Kozima and Yano, 2001, Imai et al., 2001). However, their mechanisms of joint attention were fully-developed by the designers in advance, and it was not argued how the robot can acquire such an ability of joint attention through interactions with its environment.

Butterworth and Jarrett (1991) suggested that the infant develops the ability of joint attention in three stages: ecological, geometric, and representational stages. In the first stage, the infant at the 6th month has a tendency to attend to an interesting object in

its view regardless of the caregiver's attention (see Figure 1 (a)). At the 12th month, that is the second stage, the infant begins to track the caregiver's gaze and watches the object that the caregiver attends to (see Figure 1 (b)). However, even at this stage, the infant exhibits the gaze following only when the object is within the field of the infant's view. In the final stage, the infant at the 18th month is able to turn around and attend to the object that the caregiver attends to even if the object is outside the infant's first view (see Figure 1 (c)). The developmental phenomena of the infant's joint attention can be explained in this way, however, its developmental mechanism has not been revealed yet. For this problem, Fasel *et al.* (2002) presented a developmental model of joint attention based on a proper interaction of innate motivations and contingency learning. However, the validity of their model has not been verified through the implementation to an artificial agent. Nagai *et al.* (2002) proposed a constructive model by which a robot learns joint attention through interactions with a human caregiver. They showed that the robot can acquire the ability of joint attention and the learning becomes more efficient owing to the developments of the robot's and the caregiver's internal mechanisms. However, their intention was not to explain the staged developmental process of the infant's joint attention.

This paper presents a constructive model which enables a robot to acquire the ability of joint attention without a controlled environment nor the external task evaluation and to demonstrate the staged developmental process of the infant's joint attention. The proposed model consists of the robot's embedded mechanisms: visual attention and learning with self-evaluation. The former is to find and attend to a salient object in the robot's view, and the latter is to evaluate the success of visual attention and then learn a sensorimotor coordination. Since visual attention does not always correspond to joint attention, the robot may have incorrect learning situations for joint attention as well as correct ones. However, the robot is expected to statistically lose the learning data of the incorrect ones as outliers because the object position that the robot attends to changes randomly and the data of the incorrect ones has a weaker correlation between the sensor input and the motor output than that of the correct ones. As a result, the robot acquires the appropriate sensorimotor coordination for joint attention in the correct learning situations. It is expected that the robot performs the staged developmental process of the infant's joint attention by changing the attention mechanism from the embedded one, that it visual attention, to the learned one, that is the acquired sensorimotor coordination.

The rest of the paper is organized as follows. First, how the proposed model affords the ability of joint attention based on visual attention and learning with self-evaluation is explained. Next, we describe the experiment in which the validity of the proposed model is verified. Finally, conclusion and future work are given.

## 2. The development of joint attention based on visual attention and learning with self evaluation

### *2.1 Basic idea*

An environmental setup for joint attention is shown in Figure 2, in which a robot with two cameras, a human caregiver, and multiple salient objects are indicated. The environment is not controlled, in other words, the objects are at random positions in each trial. The caregiver attends to one of the objects (in Figure 2, it attends to the square object). The robot receives the camera image $\boldsymbol{I}$ and the angles of the camera head $\boldsymbol{\theta} = [\theta_{pan},\ \theta_{tilt}]$ as inputs, and outputs the motor command to the camera head $\boldsymbol{\Delta\theta} = [\Delta\theta_{pan},\ \Delta\theta_{tilt}]$ to attend to an object. The joint attention task in this situation is defined as a process that the robot outputs the motor command $\boldsymbol{\Delta\theta}$ based on the sensor inputs $\boldsymbol{I}$ and $\boldsymbol{\theta}$, and consequently attends to the same object that the caregiver attends to.

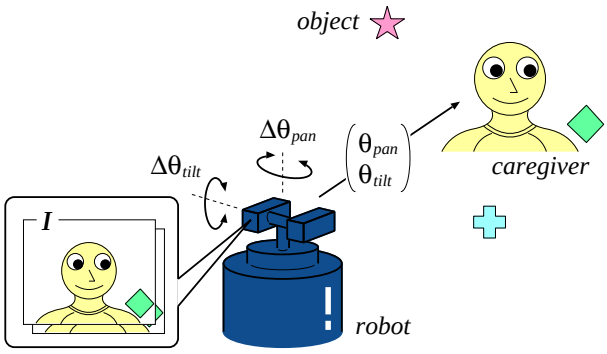



Figure 2: An environmental setup for joint attention between a robot and a human caregiver

For the development of joint attention, the robot has two embedded mechanisms:

(a) *visual attention*: to find and attend to a salient object in the robot's view, and

(b) *learning with self-evaluation*: to evaluate the success of visual attention and then learn a sensorimotor coordination.

Based on the embedded mechanisms, the robot acquires the ability of joint attention as follows. First, the robot attends to the caregiver who attends to an

object. If a salient object is observed in the robot's view, the robot shifts its gaze direction from the caregiver's face to the object based on the visual attention mechanism. When visual attention succeeds, the robot evaluates it and then learns the sensorimotor coordination between the inputs $\boldsymbol{I}$ and $\boldsymbol{\theta}$, and the output $\boldsymbol{\Delta\theta}$ based on the mechanism of learning with self-evaluation.

Since visual attention does not always correspond to joint attention, the robot may have two kinds of learning situations: correct learning situations for joint attention and incorrect ones.

- In the former case, that is when the robot attends to the same object that the caregiver attends to, the robot can acquire the appropriate sensorimotor coordination for joint attention.
- In the latter case, that is when the robot attends to the different object from that the caregiver attends to, the robot cannot find the sensorimotor correlation since it is supposed that the object position that the robot attends to changes randomly.

Therefore, the incorrect learning data would be expected to be statistically lost as outliers through the learning, and the appropriate sensorimotor correlation for joint attention survives in the learning module. Furthermore, by activating the learning module, which has already acquired the sensorimotor coordination, to attend to an object instead of the visual attention mechanism, the robot can reduce the proportion of the incorrect data and acquire more appropriate coordination for joint attention. Through the above learning process, the robot acquires the ability of joint attention without a controlled environment nor external task evaluation.

## 2.2 A constructive model for joint attention

The proposed constructive model for joint attention is shown in Figure 3. As described above, the robot receives the camera image $\boldsymbol{I}$ and the angle of the camera head $\boldsymbol{\theta}$ as the inputs and outputs the motor command to the camera head $\boldsymbol{\Delta\theta}$. The following modules corresponding to (a) visual attention and (b) learning with self-evaluation constitute the proposed model.

(a-1) *Salient feature detector* extracts distinguishing image areas from $\boldsymbol{I}$.

(a-2) *Visual feedback controller* receives the detected image features and outputs ${}^{VF}\boldsymbol{\Delta\theta}$ to attend to an interesting object.

(b-1) *Internal evaluator* drives the learning mechanism in the learning module when the robot attends to the interesting object.

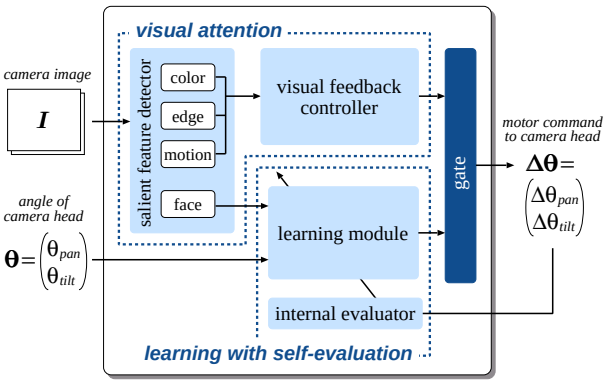



Figure 3: A constructive model for joint attention based on visual attention and self learning

(b-2) *Learning module* receives the image of the caregiver's face and $\boldsymbol{\theta}$ as the inputs and outputs ${}^{LM}\boldsymbol{\Delta\theta}$. This module learns the sensorimotor coordination when the internal evaluator triggers it.

In addition to these modules, the proposed model has another one to arbitrate the output of the robot.

(c) *Gate* makes a choice between ${}^{VF}\boldsymbol{\Delta\theta}$ and ${}^{LM}\boldsymbol{\Delta\theta}$, and outputs $\boldsymbol{\Delta\theta}$ as the robot's motor command.

The following sections explain these modules in detail.

### 2.2.1 Salient feature detector

The salient feature detector extracts distinguishing image areas in $\boldsymbol{I}$ by color, edge, motion, and face detectors. The color, edge, and motion detectors extract the objects ($i = 1, \ldots, n$) which have bright colors, complicated textures, and motions, respectively. Then, the salient feature detector selects the most interesting object $i_{trg}$ among the extracted objects by comparing the sum of the interests of all features.

$$i_{trg} = \arg\max_i (\alpha_1 f_i^{col} + \alpha_2 f_i^{edg} + \alpha_3 f_i^{mot}), \quad (1)$$

where $f_i^{col}$, $f_i^{edg}$, and $f_i^{mot}$ indicate the size of the bright color area, the complexity of the texture, and the amount of the motion, respectively. The coefficients ($\alpha_1$, $\alpha_2$, $\alpha_3$) denote the degrees of the interests of three features, that are determined according to the robot's characteristics and the context. At the same time, the face detector extracts the face-like stimuli of the caregiver. The detection of face-like stimuli is a fundamental ability for a social agent; therefore, it should be treated in the same manner as the detection of the primitive features. The detected primitive feature of the object $i_{trg}$ and the face-like one of the caregiver are sent to the visual feedback controller and the learning module, respectively.

### 2.2.2 Visual feedback controller

The visual feedback controller receives the detected image feature of the object $i_{trg}$ and outputs the motor command ${}^{VF}\boldsymbol{\Delta\theta}$ for the camera head to attend to $i_{trg}$. First, this controller calculates the object position $(x_i,\ y_i)$ in the camera image. Then, the motor command ${}^{VF}\boldsymbol{\Delta\theta}$ is generated as

$$ {}^{VF}\boldsymbol{\Delta\theta} = \begin{pmatrix} {}^{VF}\Delta\theta_{pan} \\ {}^{VF}\Delta\theta_{tilt} \end{pmatrix} = g \begin{pmatrix} x_i - cx \\ y_i - cy \end{pmatrix}, \quad (2) $$

where $g$ is a scalar gain and $(cx,\ cy)$ denotes the center position of the image. The motor command ${}^{VF}\boldsymbol{\Delta\theta}$ is sent to the gate as the output of the visual feedback controller.

As described above, visual attention that is one of the robot's embedded mechanisms is performed by the salient feature detector and the visual feedback controller.

### 2.2.3 Internal evaluator

The other embedded mechanism that is learning with self-evaluation is realized by the internal evaluator and the learning module.

The internal evaluator drives the learning mechanism in the learning module when the following condition is met:

$$ \sqrt{(x_i - cx)^2 + (y_i - cy)^2} < d_{th}, \quad (3) $$

where $d_{th}$ is a threshold for evaluating whether the robot watches an object in the center of the camera image or not. Note that the internal evaluator does not know whether joint attention has succeeded or failed but knows whether visual attention has done.

### 2.2.4 Learning module

The learning module consists of a three-layered neural network. In the forward processing, this module receives the image of the caregiver's face and the angle of the camera head $\boldsymbol{\theta}$ as inputs, and outputs ${}^{LM}\boldsymbol{\Delta\theta}$ as a motor command. The caregiver's face image is required to estimate the motor command ${}^{LM}\boldsymbol{\Delta\theta}$ to follow the caregiver's gaze direction. The angle of the camera head $\boldsymbol{\theta}$ is utilized to move the camera head incrementally because the caregiver's attention cannot be narrowed down to a particular point along the line of the caregiver's gaze. The generated motor command ${}^{LM}\boldsymbol{\Delta\theta}$ is sent to the gate as the output of the learning module.

In the learning process, this module learns sensorimotor coordination by back propagation when it is triggered by the internal evaluator. As mentioned above, the internal evaluator drives the learning module according to the success of visual attention, not joint attention, this module has correct and incorrect learning data for joint attention. In the former case, the learning module can acquire the appropriate correlation between the inputs, the caregiver's face image and $\boldsymbol{\theta}$, and the output $\boldsymbol{\Delta\theta}$. On the other hand, in the latter case, this module cannot find the appropriate sensorimotor coordination. However, the learning module is expected to statistically lose the incorrect data as outliers as described in 2.1 while the learned sensorimotor coordination of the correct data survives in the learning module. As a result, the survived correlation in the learning module allows the robot to realize joint attention.

### 2.2.5 Gate

The gate arbitrates the motor command $\boldsymbol{\Delta\theta}$ between ${}^{VF}\boldsymbol{\Delta\theta}$ from the visual feedback controller and ${}^{LM}\boldsymbol{\Delta\theta}$ from the learning module. The gate sets a gating function to define the selecting rate of the outputs. In the beginning of the learning, the selecting rate of ${}^{VF}\boldsymbol{\Delta\theta}$ is set to a high probability because the learning module has not acquired the appropriate sensorimotor coordination for joint attention yet. On the other hand, in the latter stage of the learning, the output ${}^{LM}\boldsymbol{\Delta\theta}$ from the learning module, which has acquired the sensorimotor coordination for joint attention, becomes more probable to be selected. As a result, the robot can increase the proportion of the correct learning situations according to the learning advance. It allows the learning module to acquire more appropriate sensorimotor coordination for joint attention.

## 2.3 Incremental learning

It is expected that the proposed model makes the robot acquire the ability of joint attention through the following incremental learning process.

*stage I:* In the beginning of the learning, the robot has a tendency to attend to an interesting object in the field of the robot's view based on the embedded mechanism of visual attention since the gate mainly selects ${}^{VF}\boldsymbol{\Delta\theta}$. At the top of Figure 4, the robot outputs ${}^{VF_1}\boldsymbol{\Delta\theta}$ or ${}^{VF_2}\boldsymbol{\Delta\theta}$ case by case and watches one object regardless of the direction of the caregiver's attention. At the same time, the robot begins to learn the sensorimotor coordination in each case.

*stage II:* In the middle stage of the learning, the robot is able to realize joint attention owing to the learning in *stage I* if the object that the caregiver attends to is observed in the robot's first view. At the middle left of Figure 4, the learning module has acquired the sensorimotor coordination of ${}^{LM_1}\boldsymbol{\Delta\theta}$ because only that of ${}^{VF_1}\boldsymbol{\Delta\theta}$ had the correlation in *stage I*.

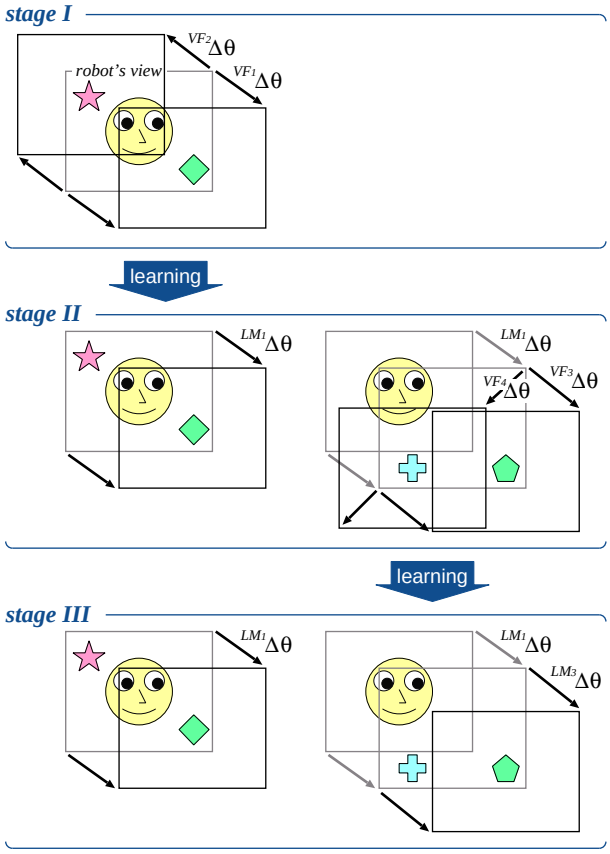



Figure 4: The incremental learning process of joint attention. The robot acquires the sensorimotor coordination of ${}^{LM_1}\boldsymbol{\Delta\theta}$ and ${}^{LM_3}\boldsymbol{\Delta\theta}$.

At the middle right of Figure 4, if the object that the caregiver attends to is out of the robot's first view, the robot can find the object not at the center but at the periphery of its view by ${}^{LM_1}\boldsymbol{\Delta\theta}$. Then, the robot outputs ${}^{VF_3}\boldsymbol{\Delta\theta}$ or ${}^{VF_4}\boldsymbol{\Delta\theta}$ to attend to an interesting object in the field of its view case by case. When visual attention succeeds, the robot learns the sensorimotor coordination in each case.

*stage III:* In the final stage, the robot has acquired the complete ability of joint attention owing to the learning in *stages I* and *II*. At the bottom of Figure 4, the robot can identify the object that the caregiver attends to by producing ${}^{LM_1}\boldsymbol{\Delta\theta}$ and ${}^{LM_3}\boldsymbol{\Delta\theta}$ even if the object is observed in the field of the robot's first view or not. The sensorimotor coordinations of ${}^{LM_1}\boldsymbol{\Delta\theta}$ and ${}^{LM_3}\boldsymbol{\Delta\theta}$ have been acquired through the learning in *stages I* and *II* because each of ${}^{VF_1}\boldsymbol{\Delta\theta}$ and ${}^{VF_3}\boldsymbol{\Delta\theta}$ had the sensorimotor correlation for joint attention

The above incremental learning process of the robot's joint attention can be regarded as equivalent to the staged developmental process of an infant's one shown in Figure 1. The *stages I, II*, and *III* of the robot correspond to the infant at the 6th, 12th, and 18th month, respectively. In addition, it is supposed in the cognitive science that the embedded mechanisms of the robot, visual attention and learning with self-evaluation, are also prepared in the infant inherently (Bremner, 1994). Therefore, the similarity of the developmental phenomena and the embedded mechanisms between the robot's joint attention and the infant's one suggests that the proposed constructive model could explain the developmental mechanism of the infant's joint attention.

## 3. Experiment

### *3.1 Experimental setup*

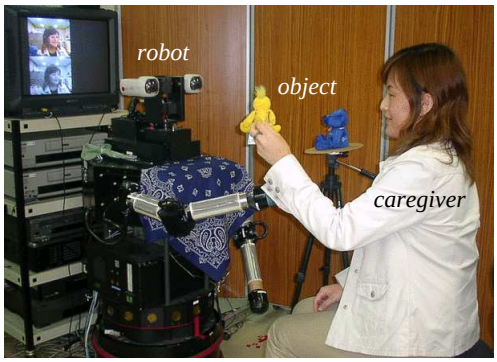
It is examined whether an actual robot can acquire the ability of joint attention based on the proposed model in uncontrolled environments including multiple salient objects. An experimental environment is shown in Figure 5 (a), and the left camera image of the robot is shown in (b). Some salient objects are places in the environment at random positions. The caregiver who sits in front of the robot attends to one object (in Figure 5, it attends to the object in its hand). The robot has two cameras and can turn them to pan and tilt simultaneously. The robot receives the camera image and detects the caregiver's face and the objects by the salient feature detector as shown in Figure 5 (b). In this experiment, the robot has the degrees of the interests of the image features $(\alpha_1,\ \alpha_2,\ \alpha_3) = (1,\ 0,\ 0)$ in Eq. (1). The threshold of the success of visual attention is defined as $d_{th}$ = (the width of the camera image)/6 in Eq. (3).

To execute the simulated learning, we acquired 125 data sets, each of which included

- the left camera image (in which the caregiver's face was extracted as a window of which size is $30 \times 25$ pixels) and the angles of the camera head (pan and tilt) when the robot attended to the caregiver's face, and

- the motor command for the camera head to shift its gaze direction from the caregiver to the object that the caregiver attended to

in advance. Then, in each trial, we took one data set from the above and placed other salient objects at random positions in the simulated environment. The number of input, hidden, and output units of the learning module were set 752 $(30 \times 25 + 2)$, 7, and 2, respectively. Under this condition, the robot repeated alternately the trial and the learning based on the proposed model.




(a) an experimental environment

(b) the left camera image of the robot (left: the detected result of the caregiver's face by template matching, right: the detected result of the bright colors)

Figure 5: An experimental setup for joint attention

### *3.2 Learning performance in uncontrolled environments*

It is verified that the proposed model enables the robot to acquire the ability of joint attention even if multiple objects are set in the environment. The gating function, that is the selecting rate of ${}^{LM}\boldsymbol{\Delta\theta}$, is defined as a sigmoid function shown in Figure 6 (a). As the result of the learning experiment, Figure 6 (b) shows the change of the success rate of joint attention in terms of the learning time, where the number of the objects is set to 1, 3, 5, or 10. Here, the number of the object 1 means that the robot has only correct learning situations in every steps. By contrast, the number 10 means that the robot can experience the correct learning situations only at 1/10 probability at the beginning of the learning. However, the robot is expected to increase the proportion of the correct ones by utilizing the learning module which has already acquired the sensorimotor coordination until that time.

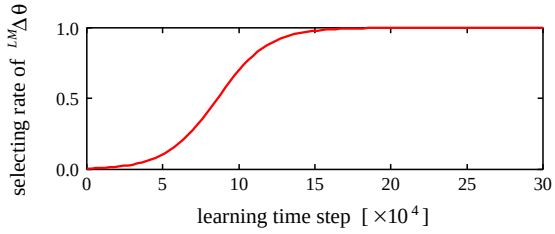
From the result of Figure 6 (b), we can see that the success rates of joint attention are at chance levels at the beginning of learning; however, they increase to high performance at the end although many objects are placed in the environment. Therefore, it can be concluded that the proposed model enables the robot to acquire the ability of joint attention without a controlled environment nor external task evaluation.

### *3.3 Incremental learning process*

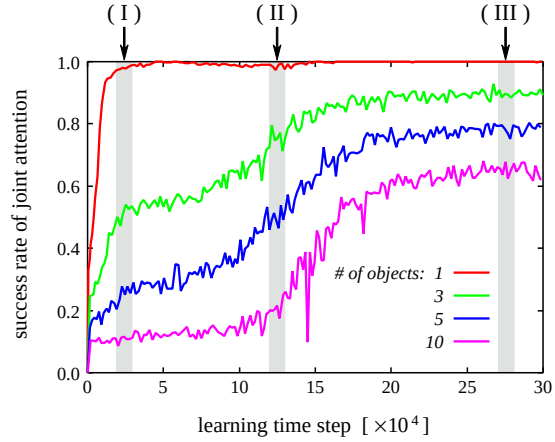
Next, we investigate the learning process of joint attention based on the proposed model. Figure 6 (c) shows the pan angle of the robot's camera head when the robot attends to an object, in which "○" and "×" indicate success of joint attention, that is the object that the robot attends to coincides with the object that the caregiver attends to, and failure, respectively. The number of objects is five, and the data are presented every 50 steps during the learning time (I) 2-3, (II) 12-13, and (III) 27-28 [$\times 10^4$], each of which is highlighted in Figure 6 (b). The pan angle is 0 [deg] when the robot attends to the caregiver, and the view range of the robot is $\pm 18$ [deg]. In other words, the objects within $\pm 18$ [deg] are observed in the field of the robot's view when the robot attends to the caregiver. From this result, we can see that the success number of joint attention increases over learning time, and at the same time, the range of the camera angle becomes wide from $\pm 18$ [deg]. These phenomena in the three stages (I), (II), and (III) can be regarded as equivalent to the infant's developmental stages of joint attention at the 6th, 12th, and 18th month shown in Figure 1. Therefore, we can conclude that the proposed model enables the robot to demonstrate the developmental process of the infant's joint attention and consequently could explain how the infant acquires the ability of joint attention.

### *3.4 Final task performance*

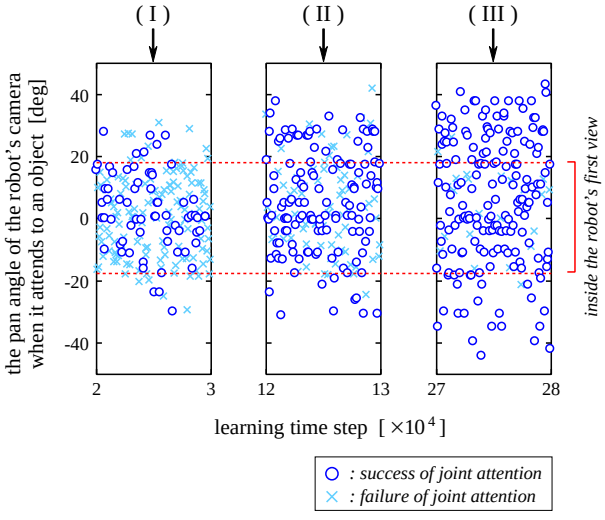
Finally, we evaluate the final task performance of the robot which has learned in the environment including five objects. Figure 7 shows the change of the robot camera image when it shifts its gaze direction from the caregiver's face to the object based on the output of the learning module. The caregiver's face image ($30 \times 25$ pixels) enclosed in a rectangle is the input of the learning module, and the straight line on the face shows the output of the learning module, in which the width and the height indicate the pan and the tilt angles of the output. The circle and the cross lines show the gazing area of the robot and the object's position, respectively. The learning module incrementally generates the motor commands ${}^{LM_1}\boldsymbol{\Delta\theta}$, ${}^{LM_2}\boldsymbol{\Delta\theta}$, and ${}^{LM_3}\boldsymbol{\Delta\theta}$ at each step, and the robot consequently attends to the object that the caregiver attends to. From this result, it is confirmed that the proposed model enables the robot to realize joint attention even if the object is far from the caregiver.




(a) gating function




(b) success rate of joint attention




(c) incremental learning process (# of objects: 5)

Figure 6: Experimental results

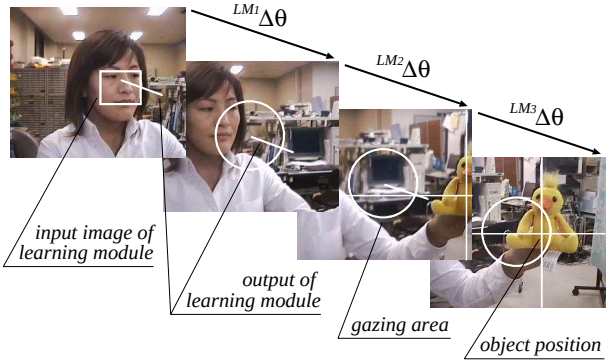



Figure 7: The change of the camera image when the robot shifts its gaze direction from the caregiver's face to the object

## 4. Conclusion

This paper has presented a constructive model which enables a robot to acquire the ability of joint attention without a controlled environment nor external task evaluation. The proposed model affords the ability of joint attention by finding the appropriate sensorimotor coordination for joint attention based on the embedded mechanisms: visual attention and learning with self-evaluation. The experimental results show that the robot acquires the ability of joint attention through the incremental learning process that is similar to the infant's developmental process of joint attention. Therefore, we can suggest that the proposed model could explain how the infant acquires the ability of joint attention.

In the future, more efficient learning mechanism should be developed so that the learning is executed not on the simulation but on the actual robot. In addition, the gating function should be designed not by the deterministic one, like a sigmoid function, but by the robot's performance of visual attention. The realization of these could make the robot become a really developmental agent. Furthermore, it would lead us to understand the mechanism of the infant's development more clearly.

## Acknowledgment

This study was performed through the Advanced and Innovational Research program in Life Sciences from the Ministry of Education, Culture, Sports, Science, and Technology, the Japanese Government.

## References

Asada, M., MacDorman, K. F., Ishiguro, H., and Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37:185–193.

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

Breazeal, C. and Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1):49–74.

Bremner, J. G. (1994). *Infancy*. Blackwell.

Brooks, R. A., Breazeal, C., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B., and Williamson, M. M. (1998). Alternative essences of intelligence. In *Proceedings of the American Association of Artificial Intelligence*, pages 961–968.

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

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development*. MIT Press.

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

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

Johnson, M. H. (1997). *Developmental Cognitive Neuroscience*. Blackwell.

Kozima, H. and Yano, H. (2001). A robot that learns to communicate with human caregivers. In *Proceedings of the First International Workshop on Epigenetic Robotics*.

Moore, C. and Dunham, P. J., (Eds.) (1995). *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates.

Nagai, Y., Asada, M., and Hosoda, K. (2002). Developmental learning model for joint attention. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 932–937.

Scaife, M. and Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253:265–266.

Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, 12:13–24.