An experiment with human-robot interaction to study intentional agency in joint visual attention

Takeshi Konno
School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1–1, Asahidai, Nomi, Ishikawa, 923–1292 Japan
Email: t-konno@jaist.ac.jp

Takashi Hashimoto
School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1–1, Asahidai, Nomi, Ishikawa, 923–1292 Japan
Email: hash@jaist.ac.jp

***Abstract*—Infants turn their own eyes to others' focus of attention. This action is called joint visual attention. It is known that the action develops from reflexive to intentional. In the early developmental process, it is pointed out that infants become intentional agents. We constructed a computational model to study intentional agency. The computational model has two main mechanisms. One is to form a memory of relationships between directions of others' gaze and objects gazed at. The other is to associate the direction of others' gaze with a target object. We suppose that the mechanisms realize an immature intentional agency. To demonstrate the mechanisms, we develop a robot to implement the computational model, and construct an experimental environment for human-robot interaction. We first test the experimental environment with the robot which produces only reflexive action. As a result, a participant in the test showed actions to explore the gaze object of the robot simply because the robot turned to face a different area from the person's gaze. From the result we noticed a problem. When the robot gazes at a different object than the one the person looked at, we cannot distinguish between two possible causes. One is a mistaken reading of the person's gaze. The other is an association with a different target object from the person's gaze. We propose a solution to solve the problem by using a measurement device for the focus of the person's gaze, and discuss a working hypothesis to demonstrate the function of our constructed mechanisms.**

I. Introduction

Joint visual attention is an action of communicative eye gaze, which is a basic ability of social communication. Joint visual attention is defined operationally as "looking where someone else is looking [2]." While this action begins as a reflexive action, then infants understand and share others' attention in terms of their intentions [4]. Further, Tomasello claims that infants first become intentional agents, before understanding others' attention [5]. However it is difficult to distinguish immature intentionality from reflexive action by observation, because intentionality is a problem concerning the infant's internal states.

A methodology called cognitive developmental robotics [1] helps us to understand subjective problems as an objective system by constructing models and operating them with robots. Nagai et al. have conducted experiments based on this approach [3]. In the research, a robot learns an action, to look in the direction of a person's gaze through visual interaction. However, the action is confined to a reflexive action, because inner growth such as intentionality does not occur in the robot.

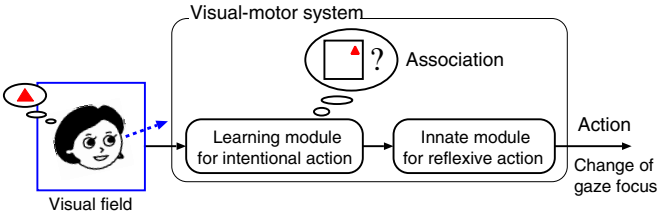

Fig. 1. System block diagram of forming intentional agency.

We have investigated the mechanisms of intentional agency through construction of a computational model which assumes formation of mental states. But we have not yet implemented the computational model to the robot. In order to demonstrate the function of the computational model, we developed a robot and started an experiment on human-robot interaction.

In this paper, first we provide a brief description of the computational model, and describe the actions which are produced by the computational model. Next, we explain an experimental environment in which people interact with the robot. Finally, we discuss our working hypothesis to demonstrate the function of our constructed mechanisms of forming intentional agency.

II. Computational model of forming intentional agency

In joint visual attention, infants perceive the parent's gaze and move their own gaze focus. To construct the computational model with mental states, internal mechanisms are important. We suppose that the formation of intentional agency can be realized by appropriately combining two modules. A learning module for intentional action is serially connected with an innate module for reflexive action (Figure 1). The learning module has two mechanisms. One is to form a memory of relationships between directions of others' gaze and objects gazed at. The other is to associate the direction of others' gaze with a target object. In our system, outputs of the learning module work as gaze targets of the system (see [6] for details). The innate module for reflexive action does not have internal states such as gaze targets.

Our system shows some different actions in exploratory activity from reflexive action. For example, if the system is in a situation with a few objects located in the same direction of the others' gaze, the system decides the target object depending

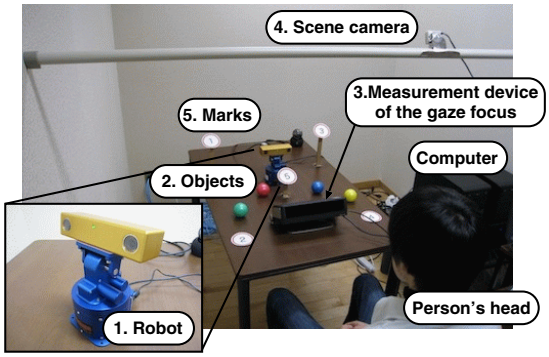

Fig. 2. An experimental environment of human-robot interaction. A person sits in front of the robot(1), and looks at the four rubber balls(2). The measurement device of the gaze focus(3), a scene camera(4), and marks(5) for calibration are located in front of the person.

on interaction with the others. If the target object is not located in the gaze area, the system explores the target object which relates to the others' gaze closely. But the system for reflexive action gazes at objects randomly after the system turns to face directions of others' gaze.

III. Human-robot interaction to demonstrate the computational model

We guess that internal mechanisms which produce internal states such as gaze targets, may let people feel intentionality. This is a function of the internal mechanism which develops to understand others' intentions and to share them. To demonstrate the function, we have started an experiment with human-robot interaction. In Figure 2, the robot is a visual-motor system which consists of a stereo camera (PGR Bumblebee2), a pan-tilt drive (TRACLabs Biclops), and a controller (personal computer). The robot measures the positions of four rubber balls (red, green, yellow, blue) and human faces by using OpenCV. A person sits in front of the robot. We use a measurement device for the focus of the person's gaze (Tobii X120).

We prepare two robots. One is to produce reflexive action (the reflexive robot, hereafter). The other is to produce intentional action (the intentional robot, hereafter). We first test the experimental environment by using the reflexive robot. A person participating in the test interacts freely with the robot. As a result, the person showed actions to explore the object of the robot's gaze simply because the robot turned to face a different area from the person. From the result we noticed a problem. When the robot gazes at a different object than the one the person looked at, we cannot distinguish between two possible causes. One is a mistaken reading of the person's gaze. The other is an association with a different target object from the person's gaze while the intentional robot interacts with the person.

Our study attempts to clarify the influence of mechanisms for intentional action. But the method to identify an object with a person's gaze has ambiguity, particularly in a situation in which a few objects are located in the same direction of the person's gaze. To solve this problem, we use a measurement device for the focus of the person's gaze. The measurement device can always measure the person's gaze focus.

Using information from the measurement device, we implement two kinds of actions at the reflexive robot. The reflexive robot always gazes either at the same object as the person (action A) or it gazes at a different object from the person's gaze with constant probability (action B). The intentional robot often gazes at a different object (action C) which is associated with the person's gaze, but this is not decided randomly.

The difference between actions A and C is remarkable. The reflexive robot and people always gaze at the same object, and the intentional robot sometimes gazes at different objects from people. In contrast, the actions B and C are only slightly different. Both robots sometimes see different objects from people. The reflexive robot gazes at objects in stochastic fluctuations, and the intentional robot decides the gaze object depending on interaction with people. We will investigate to what degree people feel the intentionality of action of each robot using some methods such as an analysis of a series of gaze objects, questionnaires, and interviews. The investigation of the differences among three actions will demonstrate the function of the constructed mechanisms.

IV. Conclusion

We study the internal states of intentional agency as a difference of mechanisms. To demonstrate the function of the mechanisms, we constructed an experimental environment for human-robot interaction. The reflexive robot gazes at objects randomly after the robot turns to face directions of person's gaze. In contrast, the intentional robot gazes at an associative object which is not determined by stochastic fluctuations, but depends on interaction with the person. Our hypothesis is that people may feel intentionality more strongly toward the intentional robot. But to demonstrate this hypothesis, the first test of experimental environment demonstrated the necessity of correct identification of the object of the person's gaze. In our experimental environment, the measurement device for the focus of the person's gaze is to be a solution.

Acknowledgements

This work was supported by Grant-in-Aid for Scientific Research (No.21700285) and Nissan Science Foundation.

References

[1] Asada,M., MacDorman,K.F., Ishiguro,H.,and Kuniyoshi,Y., "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," *Robotics and Autonomous Systems*,Vol.37,pp.185-193,(2001).

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

[3] Nagai,Y., Hosoda,K., Morita,A., and Asada,M., "A constructive model for the development of joint attention," *Connection Science*, Vol.15, No.4, pp.211-229,(2003).

[4] Tomasello,M., "Joint attention as social cognition," In Moore,C. & Dunham,P.J.(Eds.), Joint Attention: Its Origins and Role in Development, Lawrence Erlbaum, pp.103-130,(1995).

[5] Tomasello,M., "The cultural origins of human cognition," Harvard University Press, Cambridge, pp.72–73, (2000).

[6] Konno,T. and Hashimoto,T., "Developmental Construction of Intentional Agency in Communicative Eye Gaze," *Proceedings of the International Conference on Development and Learning, ICDL '06*, CD–ROM, (2006).