Towards Language Acquisition by an Attention-Sharing Robot

Hideki Kozima **Akira Ito**
Communications Research Laboratory
588-2, Iwaoka-cho, Iwaoka, Nishi-ku,
Kobe 651-2401, Japan
{xkozima,ai}@crl.go.jp

Abstract

This paper describes our preliminary research on "attention-sharing" in infants' language acquisition. Attention-sharing is the activity of paying one's attention to someone else's attentional target. This enables one to observe others' sensory-input (what they are perceiving from the target) and motor-output (what they are doing in response to the target). Being inspired by lack of attention-sharing in autistic children, we assumed that observation of others' behavior by attention-sharing plays an indispensable role in symbol acquisition. As a test-bed for attention-sharing, we are developing a robot that can follow people's attentional targets by means of monitoring their gaze-direction.

1 Introduction

Machine acquisition of natural language is one of the most challenging targets of cognitive science. As a basis for language acquisition, we deal with acquisition of a symbol system, which articulates things and events in the world into categories and gives phonological labels to the categories. The relationships between the categories and labels are arbitrary conventions shared by people, so that infants have to learn them through interaction with people.

This paper describes the role of "attention-sharing" (Baron-Cohen, 1995), especially that based on gaze, in infants' symbol acquisition. Figure 1 illustrates how attention-sharing is achieved: self (S) captures gaze-direction of an agent (A), then the self searches in the direction and identifies the target (T). Shared attention spotlights things and events being mentioned and makes the communication coherent about the same target.

2 Attention-Sharing and Symbol Acquisition

Observation of others' verbal behavior provides infants with learning data for symbol acquisition. Let us consider that an agent, looking at a cat, says "cat", as illustrated in Figure 2. In order to mimic this verbal behavior, the self has to observe the agent's sensory-input I (stimulus from the cat) and motor-output O (verbal response) and to make the association between them.

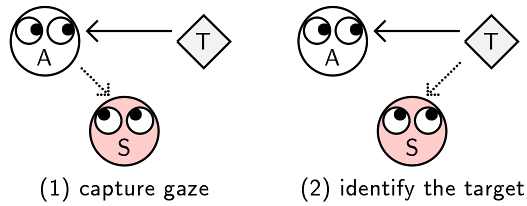


Figure 1. Attention-sharing based on gaze.

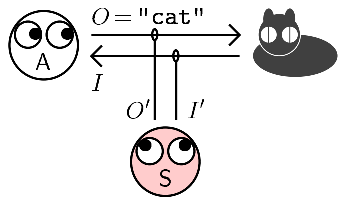


Figure 2. Observing other's verbal behavior.

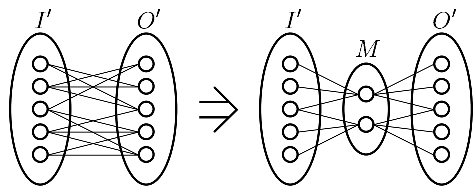


Figure 3. Introducing mediators between I/O.

Attention-sharing enables us to observe someone else's input and output, as also shown in Figure 2. Attention-sharing guarantees that I' (the self's input) resembles I, since both are paying attention to the same target. At the gaze-capturing stage (Figure 1, left), the self can observe the agent's output O and map it onto the self's motor image O'. (We assume an innate mapping mechanism like imitation of facial gestures by neonates.)

Although thus observed relationships between input space and output space may vary in many ways (size, color, tone, volume, etc.), one can construct an efficient mediator between these spaces. As illustrated in Figure 3, the complex relationships (left)

Hideki Kozima and Akira Ito (1998) **Towards Language Acquisition by an Attention-Sharing Robot.** In D.M.W. Powers (ed.) *NeMLaP3/CoNLL98: New Methods in Language Processing and Computational Natural Language Learning,* ACL, pp 245-246.

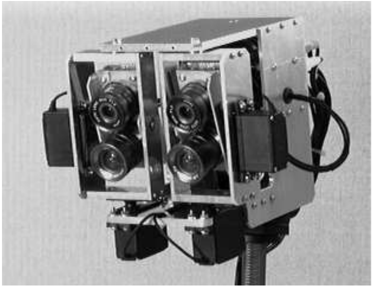


Figure 4. The attention-sharing robot.

can be decomposed into several (almost separated) components by introducing a hidden mediator space M (right), on which “symbols” can emerge.

3 Implications of Autism

Attention-sharing is commonly seen in infants at the pre-verbal stage. Its development starts before 6 months old, and is completed at around 18 months old (Butterworth, 1991). Also in some non-human primates show attention-sharing (Itakura, 1996).

Most of infants and children with autism do not show attention-sharing; being instructed by an experimenter, however, they can do it (Baron-Cohen, 1995). This means they are unaware that one’s gaze-direction implies his or her attentional target.

Being unaware of others’ attention, children with autism show typical disorders in verbal and non-verbal communication (Frith, 1989). Most of children with autism can not acquire language or use language properly. This is because (1) they failed in observing verbal (and also pragmatic) behavior of others, and (2) they failed in observing positive/negative feedback for elaborating their hypothetic language models.

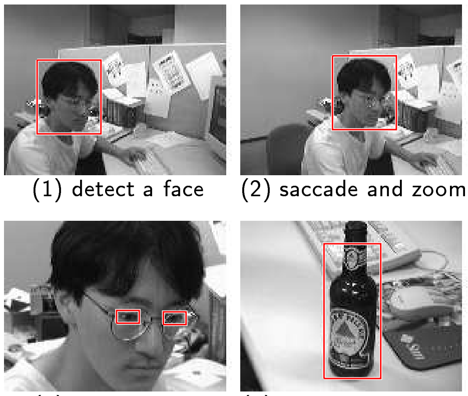
4 The Attention-Sharing Robot

We are developing a robot, *Infanoid*, as a test-bed for our model of attention-sharing. The robot is intended to create shared attention with humans in terms of monitoring their gaze-direction.

The robot has a head, as shown in Figure 4, with four CCD cameras (left/right × zoom/wide) and servo motors to drive the “eyes” at the speed of human saccade. The images taken by the cameras are sent to a workstation for gaze-monitoring.

The gaze-monitoring process consists of the following tasks, as also shown in Figure 5: (1) detect a face in a scene, (2) saccade to the face and switch to the zoom cameras, (3) detect eyes and determine the gaze-direction in terms of the position of the pupils, and (4) search for an object in the direction. If something relevant is found, the robot identifies it as the target.

We have developed a prototype of the robot and the real-time face/eye detectors. We are now working on gaze capturing and target selection. Our preliminary study found that these tasks require some top-down information like the object’s “relevance” (Sperber and Wilson, 1986) to the current context.



(1) detect a face (2) saccade and zoom
(3) capture gaze (4) identify the target

Figure 5. Gaze-monitoring process

5 Conclusion and Future Research

We described our preliminary model of attention-sharing as a device for observing learning data (others’ verbal behavior) for symbol acquisition.

The model will work in the bootstrapping stage of infants’ symbol acquisition; it only deals with referring to physical objects. Infants at this stage tend to take an unknown label as a category name of a physical object, and then apply the label to other objects with similar shape (Imai, 1997).

In future research, we have to fully implement the gaze-monitoring process and to evaluate it in human-robot interaction. Also we are planning an experiment on evaluating the accuracy of human gaze-monitoring; this will reveal how humans rely on top-down semantic/pragmatic information in attention-sharing.

References

Simon Baron-Cohen. 1995. *Mindblindness: An Essay on Autism and Theory of Mind*, MIT Press.

George Butterworth and Nicholas Jarrett. 1991. What minds have in common is space. *British Journal of Developmental Psychology*, 9:55–72.

Uta Frith. 1989. *Autism: Explaining the Enigma*, Blackwell.

Mutsumi Imai. 1997. Origins of word-learning principles. *Cognitive Studies*, 4:75–98. (*in Japanese*)

Shoji Itakura. 1996. An exploratory study of gaze-monitoring in nonhuman primates. *Japanese Psychological Research*, 38:174–180.

Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*, Blackwell.