The Role of Shared Attention in Human-Computer Conversation

Hideki Kozima and Akira Ito
Communications Research Laboratory[1]

Abstract

This paper describes our on-going project on human-computer and human-computer-human conversation. We here emphasize the role of "shared attention" in verbal and non-verbal communications. Shared attention spotlights things and events being mentioned in the conversation and makes the discourse coherent about the same topic. Being inspired by communication disorders in autism, we assume that shared attention plays an indispensable role in naturally interactive and communicative conversation, since its function is observed in infants at pre-verbal stage and its malfunction is observed in infants and children with autism. Focusing especially on "shared visual attention", that is simply "looking at the same object", we are developing a computer system that can create and maintain shared visual attention with humans by monitoring their gaze-direction. We are planning experiments on human interaction with this system in order to evaluate and elaborate our model of shared attention in conversation.

1. Introduction

Human-computer conversation is one of the most challenging targets of computational linguistics. Researchers have developed a number of computational devices for analyzing and generating speech, sentences, and discourse. Integrating these devices, however, no one achieved natural human-computer conversation like that of HAL 9000.

We emphasize here that "shared attention" (Baron-Cohen 1995) have been missing in the former studies. Shared attention, that is attention shared by speakers and hearers, conveys a clue to what has "relevance" (Sperber 1986) to the current context. It plays a dominant role not only in encoding and decoding referring expressions but also in making coherent discourse "being about what we are paying attention to". (Note that psychologists often use the term "joint attention" instead of "shared attention".)

This paper describes our on-going research on the role of shared attention in verbal and non-verbal communications. We currently deal with "shared visual attention" (Baron-Cohen 1995), that is simply "looking at the same objects", as one of the most fundamental devices of pre-verbal communication of infants as well as verbal communication of adults.

The following section briefly describes the nature of shared attention in human

[1]Iwaoka 588-2, Iwaoka-cho, Nishi-ku, Kobe 651-24, Japan. ({xkozima,ai}@crl.go.jp)

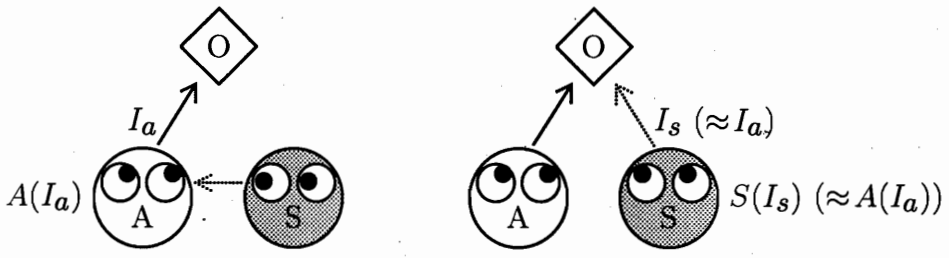


Fig. 1 Shared attention and estimation of others' mental states.

communication. Shared attention enables us to estimate others' intentions and emotions behind their explicit behavior. Section 3 introduces the attention-sharing system being developed. The system is intended to create and maintain shared visual attention by monitoring people's gaze-direction. Section 4 gives our preliminary conclusion and draws our plan for future research.

2. Shared Attention in Communication

Human communication is an activity of creating "shared world" with others. Shared world consists of things and events that are physically or mentally manipulatable from speakers and hearers, and it works as a field for exchanging information with others and understanding others' mental states.

Shared attention, especially shared visual attention, plays an indispensable role in creating shared world with others, since one's attentional target is closely related to his or her belief and desire (Frith 1991, Baron-Cohen 1995). Figure 1 illustrates how shared visual attention is created. First, Self (e.g. an infant) captures the gaze-direction of Agent (e.g. his or her caretaker). Then, Self searches in the direction and identifies Object to which Agent is paying attention.

Shared visual attention is one of the developmentally fundamental devices for communication. Visual attention-sharing is observed in infants at the pre-verbal stage: its development starts before 6 months old and completed around 18 months old (Butterworth 1991). In addition, it is also observed in some species of non-human primates, e.g. chimpanzees and orangutans (Itakura 1996).

Most infants and children with autism can not create shared visual attention with others; 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. Unawareness of others' attention results in "mindblindness" (Baron-Cohen 1995), that is a disability of estimating others' mental states in terms of shared attention, which causes autism's typical disorders in verbal and non-verbal communications.

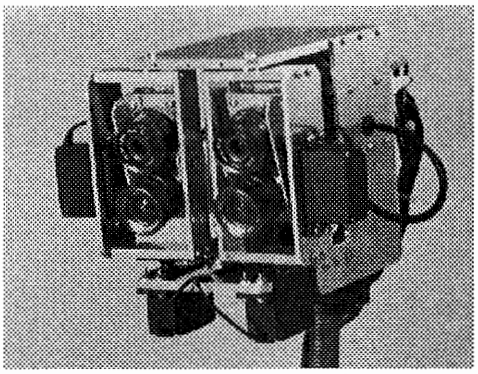


Fig. 2 Attention-sharing system.

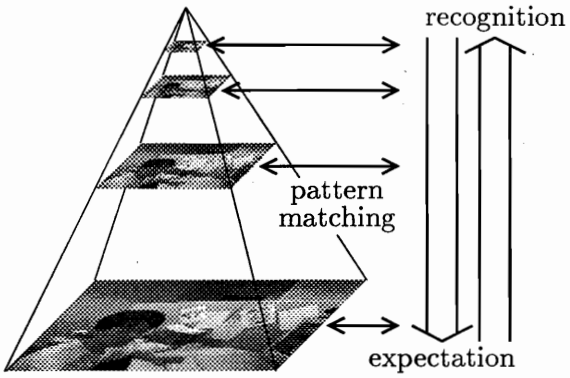


Fig. 3 Hierarchical image processing.

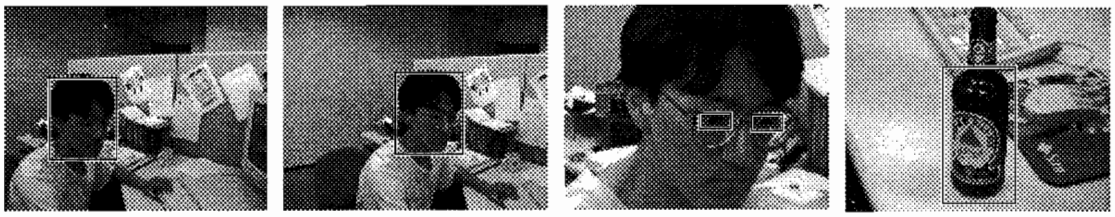
The role of shared attention in human-computer conversation lies in (1) capturing "relevance" and (2) estimating "others' mental states". First, shared attention enables us to select a target object relevant to the current context. Communication of intentions is often achieved by just pointing to a relevant object (e.g. pointing to a clock when you want to make someone hurry). Though it can tell us only physical targets in our sight, shared visual attention provides developmental basis for higher-level attention-sharing.

Secondly, shared attention enables us to estimate "others' mental states": once a hearer selected a relevant object in terms of shared attention, the hearer can co-observe speaker's sensory-input (what Agent is perceiving from the target object) and the subsequent mental states (speaker's intentions and emotions). As illustrated in Fig. 1 again, Self can estimate Agent's sensory-input I_a being perceived from Object O, then Self can simulate Agent's mental states $A(I_a)$ by applying Self's mind $S(\cdot)$ to the pseudo-input I_s thus co-observed. Since shared attention guarantees $I_a \approx I_s$ and our innate and/or cultural bias expects $A(\cdot) \approx S(\cdot)$, the simulation result $S(I_s)$ becomes a good approximation of $A(I_a)$.

3. The Attention-Sharing System

We are developing a computer system that can share visual attention with people in terms of monitoring their gaze-direction (Tanenhaus 1996). The system, though it is still under development, is intended as a computational model of the cognitive module for visual Attention-sharing that will be incorporated into our system of human-computer conversation.

The attention-sharing system consists of a robot head with anthropomorphic shape shown in Fig. 2, and a standard workstation. The robot head has four CCD monochrome cameras (left/right × zoom/wide) and four servo motors to drive the directions of the left/right "eyes" at the speed of human saccade. The images taken by these cameras are sent to the workstation for a gaze-monitoring procedure. For real-time gaze-monitoring,



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

Fig. 4 Gaze-monitoring process.

we employed "hierarchical image processing" illustrated in Fig. 3.

The gaze-monitoring procedure consists of the following computational stages on "images" and "relevance". (See also Fig. 4.)

1. Detect a face (under varying pose and size) in a complex scene.
2. Saccade to the face and switch to the zoom cameras for precise face images.
3. Detect eyes and capture the gaze-direction in terms of the position of pupils. If it is impossible, capture face-direction instead.
4. Search for an object in the gaze-direction. If something relevant to the current context is found, identify it as a target object.

We have developed a prototype of the robot head, its control module, and the real-time face/eyes detection procedure; we are now working on capturing gaze-direction and target selection. The search/identify stage requires to evaluate objects' relevance to the context. This is because human gaze-monitoring is not so precise — though we have not done the evaluation of the precision — that they would rely on semantic and pragmatic clues like relevance.

4. Conclusion and Future Research

We outlined our on-going research on the role of shared attention. Human communication is an activity of sharing one's mental state with others. Shared attention plays an indispensable role in (1) selecting a target object relevant to the current context and (2) estimating others' mental states produced by the target object.

We are developing an attention-sharing system which can create and maintain shared visual attention with people in terms of monitoring their gaze-direction. We have achieved the real-time face/eyes detection mainly by a bottom-up approach; we found that capturing gaze-direction and selecting a target require a top-down approach, namely evaluation of objects' relevance to the current context.

Our short-term goal is to complete the gaze-monitoring process, evaluate it in human

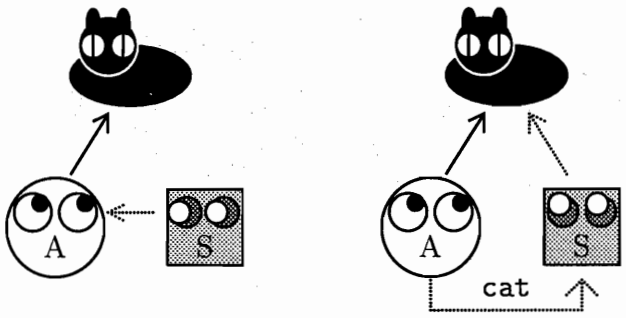


Fig. 5 Language acquisition by attention-sharing.

experiments, and elaborate our model of shared attention. Also we are planning an experiment on evaluating human gaze-monitoring precision. This will reveal how humans rely on top-down semantic and pragmatic expectations in gaze-monitoring.

We have two long-term goals. One is to incorporate the gaze-monitoring system into human-computer conversation systems. For this, the system has to extract linguistic context from the discourse in order to evaluate objects' relevance to the context. The other is to construct a model of infants' acquisition of the symbolic system of the first language. A symbolic system is a set of arbitrary associations between expressions and meaning; it articulates things and events in the world into categories and gives phonological representations to the categories. Attention-sharing with caretakers, as is illustrated in Fig. 5, will enable infants to learn associations between representations (e.g. caretakers' utterances) and meanings (e.g. estimated caretakers' mental states produced by target objects).

References

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

Butterworth, G. and Jarrett, N.: What minds have in common in space: spatial mechanisms serving joint visual attention in infancy, *British Journal of Developmental Psychology*, Vol.9, pp.55-72, 1991.

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

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

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

Tanenhaus, M. K. and Spivey-Knowlton, M. J.: Eye-tracking, *Language and Cognitive Processes*, Vol.11, pp.584–588, 1996.