Self-observation Principle for Estimating Peers' Internal State
– New Computational Theory on Communication –

MAKINO, Takaki and AIHARA, Kazuyuki
(Institute of Industrial Science, Tokyo University)

We propose a computational theory on estimating the internal states of peers, which is the basis of information processing in pre-verbal communication.

The brain mechanism behind human communication is an important theme in cognitive sciences. Communication does not depend solely on linguistic abilities, as demonstrated by the fact that a three-month-old baby has a natural ability to communicate with his/her parents. Even communication between monkeys involves the emergence of new movements, such as tactics and cooperation. Although linguistic studies implicitly presuppose such preverbal communication, studies on the ability itself have tended to focus on phenomenal and descriptive research. Consequently, the mechanisms behind preverbal communication are still poorly understood.

One way of solving such a large, complex problem is through the constructive approach, i.e., understanding the brain by constructing. Given a goal function, we can look for algorithms suitable for this, without referring to the human brain. However, to study human communication with this approach, we need a concrete *computational theory* on communication [1]. Since previous studies on constructive approach (e.g. [2]) lacks computational theory, it is hard to evaluate and extend the studies.

In this presentation, we investigate the core computation in communication, estimating peers' internal states, within the framework of estimating dynamics. Although this framework is a general mathematical framework that has no direct relation to human beings, its application reveals the computational tasks involved in human communication

We then point out two difficulties, which prevent the theory from being applied in practice. The first is limitations in the parameter dimensions of the estimator. To estimate another's internal states only by observation, the estimator has to have significantly larger parameter dimensions than the target. Such limits prevent us from estimating peers, who are equivalent to the estimator. The second is conversion from objective to subjective information and vice versa. Even if an estimator has reconstructed the other's dynamics model in himself/herself, the input to the model is subjective information from the viewpoint of the other. To use the model, the estimator has to convert his/her observation into the target's subjective information.

To solve these problems, we propose a new computational theory based on the *self-observation principle* [3], where the dynamics model, which is learned through objective observation of the self, is applied to others to estimate their internal states. The first difficulty is reduced by prior knowledge on the dynamics of others provided by learning one's own dynamics, while the second is resolved through learning the association between one's own subjective state and objective self-observation.

We also discuss the proposed theory within the context of related scientific domains, and explain why the self-learning dynamics model is related to human self-consciousness and mirror neurons.

[1] Marr D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco.
[2] Wolpert DM, Doya K, Kawato M. 2003. A unifying computational framework for motor control and social interaction. *Phil. Trans. R. Soc. Lond. B*, **358**, 593–602.
[3] Takaki Makino and Kazuyuki Aihara. 2003. Self-observation principle for estimating the other's internal state - a new computational theory of communication. Mathematical Engineering Technical Reports METR 2003-36, Department of Mathematical Informatics, the University of Tokyo.