Joint Attention and Dynamics Repertoire in Coupled Dynamical Recognizers

Takashi Ikegami and Hiroyuki Iizuka
Department of General Systems Sciences,
Graduate School of Arts and Sciences,
University of Tokyo
3-8-1 Komaba, Tokyo 153-8902
{ikeg,ezca}@sacral.c.u-tokyo.ac.jp

Abstract

A coupled dynamical recognizer is proposed as a model for simulating turn-taking behavior. An agent is modeled as a mobile robot with two wheels. A recurrent neural network is used to produce the motor outputs. By controlling this, agents compete to take turns on a two dimensional arena. By using the genetic algorithm technique, we show that turn-taking behavior is developed between two agents. It is worth noting that turn-taking is established with a variety of dynamics. A coupling between agents is sensitive to the dynamics and we discuss the sensitivity by referring to Trevarthen's double monitor experiments.

1 Intersubjectivity and Joint Attention

Here in this paper, we propose a simulation study of joint attention via coupled dynamical recognizers. There are many ways to understand psychological phenomena not directly by studying human behavior but by computer simulations and robot experiments (e.g. B.Scassellati (1999), K. Dautenhahn (1999)). To bridge between simulation studies and psychology, we think it worth discussing some key concepts such as joint attention and intersubjectivity as a Rosetta stone.

1)Joint attention:

Joint attention requires a coordinated preverbal behavior among 2 or more persons. A simple example is a children's pointing under the attention of the mother. It is a process of sharing one's experience of observing an object or events with others by following pointing gestures or eye gazing.

We distinguish two types of joint attention (R.Uno and T.Ikegami (2002)). If a person uses joint attention as a tool to achieve a goal (e.g. establishing the joint attention to let your dog pick up a ball), we call it "instrumental joint attention". But if a person takes joint attention itself as a goal, we call it "participatory joint attention". For example, two looking at the same sunset establish the participatory joint attention as it doesn't require further achievements.

2)Intersubjectivity:

Intersubjectivity is to share mental states and intentions with others. Trevarthen , who first noticed its importance, classified two types of intersubjectivity. A primary intersubjectivity is found in between an infant and a caretaker. It is thought that there exists an innate ability to orangize intersubjectivity. From around 7 months old, a secondary intersubjectivity develops. It starts to involve a third object or event besides person-person relationship. This secondary intersubjectivity is thus close to instrumental joint attention in many ways.

It is clear from the above definitions that intersubjectivity is a close concept of joint attention. Here we realize that to study turn-taking as objective realization of those concepts. We see that intersubjectivity can be taken as a form of interaction between subjects without explicitly referring to sensory-motor coordination. On the other hand, turn-taking behavior is often taken as a mere sensory-motor coordination between subjects without referring to its intersubjectivity. A glue between intersubjectivity and turn-taking is given by joint attention. In particular, behavior of participatory joint attention has both aspects of sensory-motor coordination and intersubjectivity. We also think that the synchrony in preverbal interaction (i.e. turn-taking) triggers the emergence of intersubjectivity via joint attention.

In order to clarify the interrelation between joint attention, intersubjectivity and turn-takings, we introduce another notion, i.e. novelty. Joint attention is, not always but, something to do with surprise and expectation. A kind of "novelty" attracts other person to engage in communication. It is known and used in developmental studies that infants watch longer at the new/surprising events (M.Legerstee (2001)). Namely, novelty also attracts infants to the new event and thus interaction between infants and event is established. We use such novelty as unpredictability as a necessary factor to start and maintain interaction.

A novelty=unpredictability should be continuously generated in order to sustain interactions but at the same time it is also true that novelty cannot be prepared beforehand and too much novelty breaks up the interaction. As such joint attention means to continuously introduce novelty into the interacting field.

In the following, we introduce a basic framework called coupled dynamical recognizers and showing complex dynamics underlies turn-taking.

2 Coupled Dynamical Recognizers

Neural networks with recurrent interactions have been used to imitate human language performance (J.L. Elman et al. (1996)), to mimic finite automaton behavior (J.B.Pollack (1991)), and to manipulate robot navigation (J.Tani (1998)). Extending Pollack's terminology, we term this a dynamical recognizer in the sense that a cognitive agent should be a dynamical agent, i.e., a momentary agent that is always changing its internal state. As Pollack first showed explicitly, a dynamical recognizer (DR) can imitate the behavior of a finite automaton. By using a back propagation (through time) algorithm, we can train a dynamical recognizer to learn context sensitive grammar (A. Blair and J.B. Pollack (1997)).

To see how well a dynamical recognizer learns to imitate a given automaton, we often examine geometrical patterns in the 'context space', i.e., a plot of output context node states against a series of random inputs. If a dynamical recognizer can successfully imitate the given finite automaton, the context space plotting will show finite islands of clusters, and a clear correspondence is observed between each cluster and each node of the finite automaton. When the dynamical recognizer fails to imitate or the opponent is not a finite automaton, the context space plotting shows a stretched and folded, fractal-like structure.

By coupling two DRs, we have simulated the iterated prisoner's dilemma game (T.Ikegami and M.Taiji (1998); M.Taiji and T.Ikegami (1999)), Dubey's game (T.Ikegami and M.Taiji (1999)),coalition game (G.Morimoto (M.Taiji and T.Ikegami)) and discourse interactions (I.Igari and T.Ikegami (2001)). In the present study, we show how turn taking behavior is organized by a coupled DRs.

3 The Model

Each agent here has a circular body of radius R, with two diametrically opposed motors (Fig. 1). The motors can take the agent backwards and forward in a 2-D unstructured and unlimited arena, which is described by the following equation of motion of an agent's heading angle (θ) and the velocity (v) to the direction:

$$M\dot{v} + D_1 v + f_1 + f_2 = 0, \quad (1)$$
$$I\ddot{\theta} + D_2\dot{\theta} + \tau(f_1, f_2) = 0, \quad (2)$$

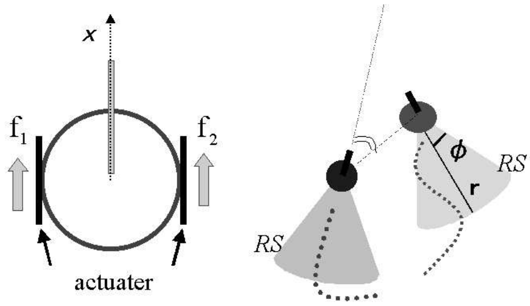

Figure 1: Left: a schematic view of the mobile robot with two wheels (actuators). It computes the forward force vector and the torque strength from the force vector (f_1, f_2) on each actuator. Right: Two mobile robots interact to perform turn-taking behaviour by sensing each other's position, distance and angle. It is called a robot A's turn when A enters B's rear side (RS) position. The shape of this RS is parameterized by r and ϕ.

where f_1 and f_2 are the forward driving force, and τ gives the torque. D_1 and D_2 express the resistance coefficients. Also the agents have mass (M) and the momentum (I). We iterate the equations using the Runge-Kutta method. At each time step, agents compute the forces from inputs by using the internal DR that will be explained in below.

We assume no collision between agents because we mainly focus on the internal states of agents that generate turn-taking. Two agents try to coordinate turn-taking behavior; mutually getting behind the other. Since they cannot get behind each other simultaneously, the turn-taking is not achieved if both agents play chaser. Of course, if both agents play evader, mutual turn taking cannot be achieved neither. Therefore, it is necessary to have spontaneous symmetry break down that one plays a role of chaser and the other plays a role of evader. But the mere symmetry breaking is not sufficient. Temporal role changing is also required. By using coupled DRs, we focus on how turn taking dynamics self-organizes.

3.1 Agents

We designed the agents as a dynamical recognizer (i.e. equipped with recurrent neural networks). Inputs to an agent are the other agent's position, distance and heading angle, relative to the agent. They move freely in the arena by using two motors' outputs, which are computed at each time-step. The agent predicts the other one's next relative position. They are assigned at three output neurons. The dynamics of the network are expressed by the following equations at the game time-step t,

$$h_j(t) = g(\sum_i w_{ij} y_i(t) + \sum_l w'_{lj} c_l(t-1) + b_{j1}), \quad (3)$$
$$z_k(t) = g(\sum_j u_{jk} h_j(t) + b_{j2}), \quad (4)$$

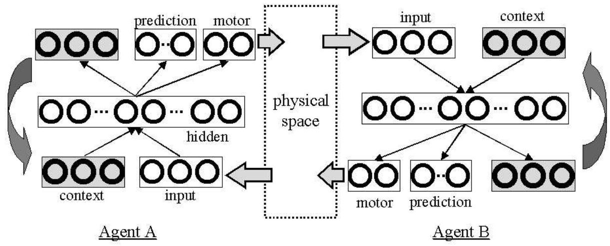

Figure 2: Coupled dynamical recognizers with three layers. Input nodes receive the other agent's relative position. The last layer consists of three kinds of nodes: context, prediction and motor. Context nodes feed back to the input layer. Prediction nodes output the other's relative position in the next time step. Motor nodes provide the force vector (f_1 and f_2).

$$c_l(t) = g(\sum_l u'_{jl} h_j(t) + b_{j3}), \quad (5)$$

$$g(x) = 1/(1+\exp^{-x}), \quad (6)$$

where y_i, z_k, h_j and c_l represent input, output, hidden and context nodes, respectively. The respective number of nodes in these layers is set to $(I, K, J, L) = (3, 5, 10, 3)$ throughout this paper. The symbols w_{ij}, u_{jk}, w'_{lj} and u'_{jl} denote the weights from input to hidden, hidden to output, context to hidden and hidden to context neurons, respectively and a parameter b gives a bias node. This network architecture is evolved via a genetic algorithm, which will be explained in the next section.

4 Evolutionary design of neural architecture

We update the weights according to turn-taking performance. In practice, the weight set of the neural networks has a vector representation of the real weight values, which should be evolved through GA.

There are two populations which separately evolved by GA. Each population contains P individuals. The performance of the any paired agents from the separated populations are evaluated at each GA generation. Agents which can equally exchange turns are evaluated to have a higher fitness. At first, individuals in each population is initialized with randomized weight values. Then, we calculate the fitness of each individual based on this performance.

The highest value is given when both agents take their turn alternately and the agents can predict each other's behaviour. A one-sided (i.e. role-fixed) behaviour is associated with the lower fitness values. Practically, the fitness of an agent a from a population (A) against an agent b from the other population (B) is calculated as follows. In below, we define a total fitness F as a sum of two fitness associated with prediction and turn-taking, respectively. When an agent get behind the other agent is defined as its turn and the rear scope is specified as RS, that is parameterized by two parameters r and ϕ (see Fig. 1). The agent in this scope is said to be in his turn and being rewarded. A spatial position of the b-th agent at time step t is represented by $Pos_b(t)$. This is compared with the a-th agent's prediction value $Pos_{a\to b}$. Thus the squared difference (Eq.(11)) evaluates the precision of the a-th agent's prediction.

$$F_a = s_1 \times F_a^{turn} + s_2 \times F_a^{predict}, \quad (7)$$

$$F_a^{turn} = \frac{1}{P}\sum^{P}\left(\sum_t^T g_a(t) \times \sum_t^T g_b(t)\right), \quad (8)$$

$$g_a(t) = \left\{ \begin{array}{ll} 1 & Pos_a(t) \in RS_b(t) \\ 0 & Pos_a(t) \notin RS_b(t) \end{array} \right\}, \quad (9)$$

$$F_a^{predict} = -\frac{1}{P}\sum^{P}\left(\sum_t^T P_a(t) \times \sum_t^T P_b(t)\right), \quad (10)$$

$$P_a(t) = (Pos_b(t) - Pos_{a\to b}(t))^2, \quad (11)$$

The performance of turn-taking is evaluated for different time lengths ($T = 500, 1000$ and 1500), so that agents can't tell when the evaluation time is over. Evaluating the turn-taking performance at each GA generation, we leave the best E individuals in each population and let them reproduce with certain mutation rates. The GA generation proceeds by repeating this procedure and the DR structures are updated. We use $P = 15$ individuals and select $E = 5$ best agents for each GA generation.

It should be noted here that the sensory noises are added to input neurons during each run. Therefore agents have to take turns under the noisy environment.

5 Simulation results

After several thousand GA generations, turn-taking is established between two agents. A basic dynamics of the turn-taking was observed as follows. Two agents are tuning their velocities and make turns automatically to switch from an role of evader to of chaser and *vice versa*. Some examples of spatial patterns of turn-taking are depicted in Fig.3. Here for the clarity reason, a single agent's trail is depicted.

With respect to the prediction capabilities, we notice that both agents' predictions break down when exchanging turns, while prediction holds for each turn holding phase (see Fig.4).

We classified the turn-taking behaviours roughly into 2 types; regular and geometrical ones and chaotic ones. Agents take turns at the same locations in the geometric case, but agents chaotically change when and where to exchange turns in the chaotic case (see the details in H.Iizuka and T.Ikegami (2003)).

We here introduce two "psychological " experiments to those evolved agents. Agents are sensitive to the microstructure of dynamics underlying turn-taking behaviors. This is analyzed by the following experiments.

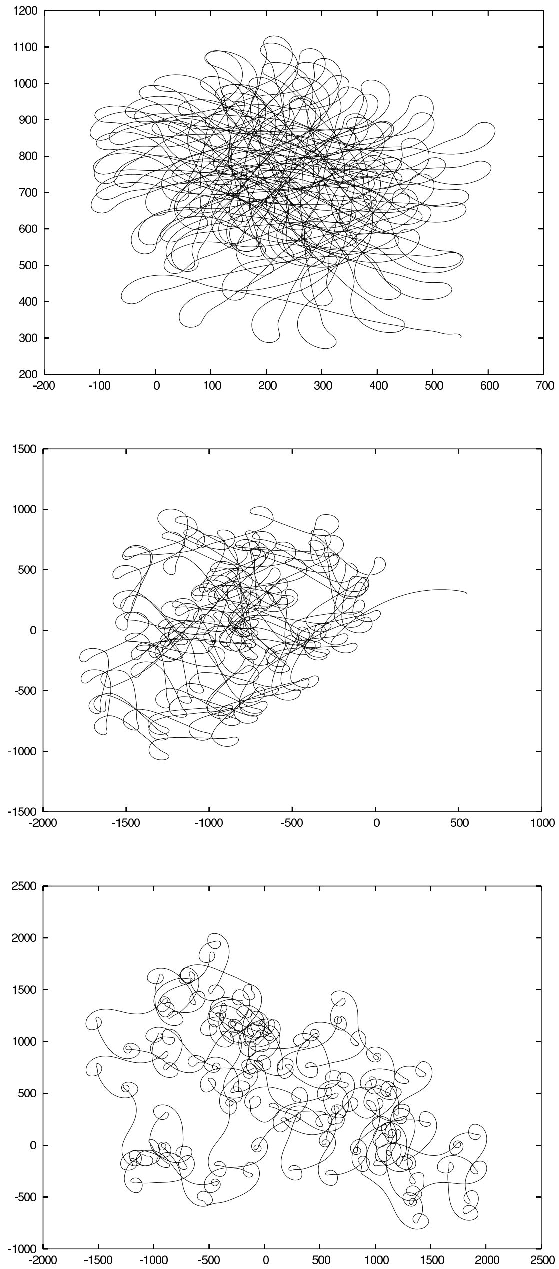

Figure 3: Examples of spatial trails on the 2-dimensional arena from 7000 (top), 8000(middle) and 10000(bottom) GA generations. Since both agents have similar trails, only a single agent's trail is drawn here. Generally, loops are corresponding to exchainging turns. All these patterns are classified as chaotic turn-takings.

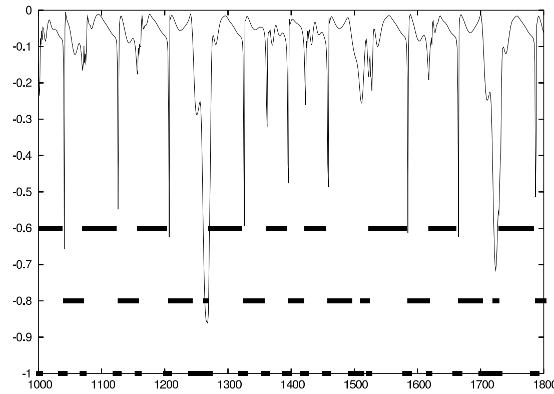

Figure 4: Prediction (top spiky lines) and turns (line segments) are drawn for an agent from 10000 GA generations. A horizontal line expresses time steps for two agents moving on the 2-dimensional arena. The top two line segments correspond to turns of the coupled agents. The bottom segments correspond to when no one gets turns. It shows that the prediction precision sharply drops off when a turn is switched.

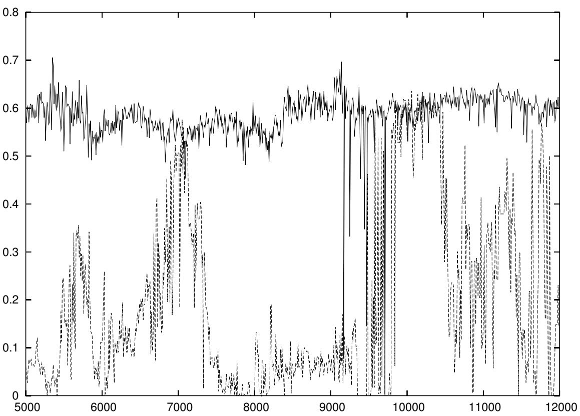

Figure 5: A best agent at the 10000 GA generation is examined to couple with the other best agent at different GA generations. The performance of turn-taking of newly coupled pair is evaluated for each generation. Generally the performance becomes lower (a dashed line) than the original performance (a solid line) of the best pair at each generation.

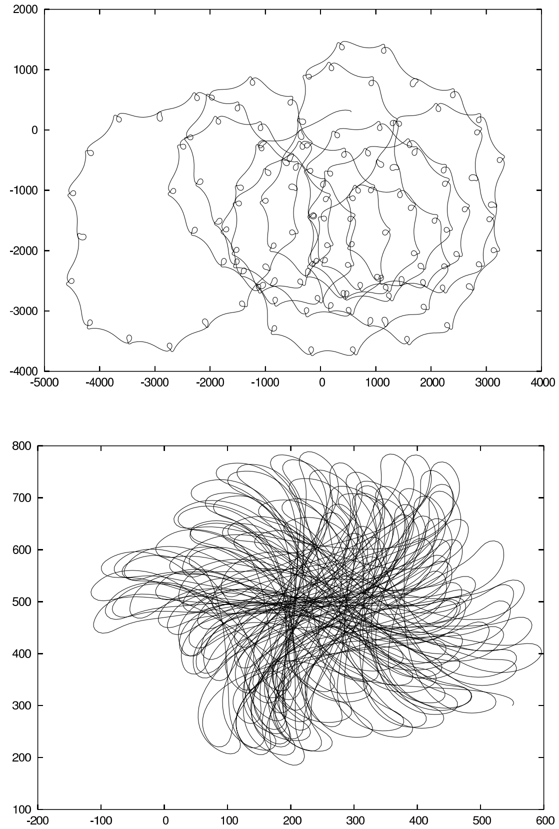

Figure 6: Spatial trails of newly coupled agents of 8000 and 10000 GA generation (upper picture) and of 7000 and 10000 GA generation (lower picture). The lower picture is similar to the one generated by the original agents of 7000 generation. However, the upper picture shows a new trail. The lower picture has much better performance than the upper picture.

(1) Coupling between two best agents from different GA steps:

After sufficient turn-taking behavior has been established between agents, we pick up two individuals from the different generation to make a novel pair. This is to examine whether the novel pair can perform trun-taking or not. The novel pairs generally show far worse performance than the original pairs (Fig.5). Within a new pair, responses of one agent to the other is often inappropriate in time. Contrary to the original pairs that show complete synchronization of turn-taking. It is worth noting that new pairs sometimes generate new spatial patterns but sometimes either dynamics takes over the other. For example in Fig 6, the agent from the 7000 GA steps can entrain the dyanmics of the agent from 10000 steps. On the other hand, the agent from the 8000 step with the agent from 10000 steps generates a new pattern. Both spatial patterns are belonging to chaotic cases.

(2) Coupling with a virtual agent:

Agents establish turn-taking behavior under the noisy environment. But they can also perform without noise (with some exceptions). We first simulate a normal turn-taking between two agents. Second, we simulate turn-taking between an agent and a recorded spatial trail generated by the other agent in the first run. The recorded trail is named a virtual agent. We perturb the virtual agent by adding a small noise on a portion of the recorded trail. In a geometrical case, a turn-taking is possible even with the perturbed virtual agent. But in a chaotic case, the small difference is amplified and a turn-taking fails to establish. Regarding that turn-taking is possible even under the noisy environment, we insist that turn-taking by the agents of chaotic case distinguishes between noisy inputs and noisy but adaptive inputs generated by the other agent.

6 Discussions

We compare this simulation result with Trevarthen's double monitor experiments between a baby and the mother (C.Trevarthen (1993)). It is reported that the mother's response to a baby should be of the right style and the right timing. Otherwise, the baby doesn't engage with her. For example, if the recorded video of the mother was displayed to the baby, the baby becomes withdrawn and depressed. This corresponds to our experiment (2) of turn-taking between an agent and a virtual one. We showed that agents become sensitive to the micro-structures of the dynamics of the other agent (e.g. when to make turns, how fast turn-taking can proceed etc.).

Turn-taking is potentially maintained by various types of dynamics as we see in our experiments. Trevarthen also reported (1977) that each infant-mother pairs developed a different style of mutual activity (e.g. forms of actions and facial expressions). In our experiment (1), a same performance allows many different style of turn-taking behaviours. More interestingly, newly coupled pairs can generate novel turn-taking dynamics. This potential diversity of dynamics repertoire is important for developing sophisticated communication such as language.

The present simulation shows how our communicaiton system works based on the fragile and unstable dynamics. Joint attention is a style of interaction between cognitive autonomous agents. It has to maintain the interaction under the uncertainty caused by the autonomy of agents. But paradoxically, joint attention works well and can transfer subtle differences of meanings between agents. All these interesting aspects of joint attention are prepared in the primitive form of interaction, i.e. turn-takings. Details of this study will be reported elsewhere (see, e.g. H.Iizuka and T.Ikegami (2002, 2003)).

Acknowledgments: This work is partially supported by Grant-in aid (No. 09640454 and No. 13-10950) from the Ministry of Education, Science, Sports and Culture.

References

A.Blair and J.B. Pollack (1993). Analysis of dynamical recognizers. *Neural Computation 9(5): 1127-1142*.

K.Dautenhahn (1999). Embodiment and interaction in socially intelligent life-like agents. *Springer Lecture Notes in Artificial Intelligence, Volume 1562, Springer, pp.102–142*.

J.L.Elman, E.A.Bates, M.H.Johnson, A.Karmiloff-Smith, D.Parisi, and K.Plunkett (1996). Rethinking innateness. *The MIT press*.

I.Igari and T.Ikegami (2001). Coevolution of mind and language. *(presented at the int'l conf. on Evolution of Language. (Paris, April, 2000) (unpublished). I.Igari, Ph.D. thesis (University of Tokyo).*

H.Iizuka and T.Ikegami (2002). Simulating Turn Taking Behaviors with Coupled Dynamical Recognizers. *the proceeding of Artificial life 8 pp. 319–328(eds. Standish et al. MIT press).*

H.Iizuka and T.Ikegami (2003). Adaptive Coupling and Intersubjectivity in Simulated Turn-taking Behaviors. *submitted to the European Conference for Artificial Life 2003*.

T.Ikegami and M.Taiji (1998). Structures of possible worlds in a game of players with internal models. *Acta Polytechnica Scandinavica Ma. 91, pp.283–292*.

T.Ikegami and M.Taiji (1999). Imitation and cooperation in coupled dynamical recognizers. *Advances in Artificial Life of Springer-Verlag, pp545–554*.

M. Legerstee. (2001) Six-Month-Old Reply on Explanatory Inference When Relating Communication to People and Manipulatory Actions to Inanimate Objects: Reply on Gergely. *Developmental Psychology, 37,5, pp.583-586.*

T.Ikegami and G. Morimoto, Chaotic Itinerancy in Coupled Dyamical Recognizers (preprint)

J.B.Pollack (1991). The induction of dynamical recognizers *Machine Learning, 7, pp.227–252*.

B.Scassellati (1999). Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. *Computation for Metaphors, Analogy and Agents, Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag*.

M.Taiji and T.Ikegami (1999). Dynamics of internal models in game players. *Physica D, 134, pp.253–266*.

J. Tani (1998) An Interpretation of the 'Self' from the Dynamical Systems Perspective: A Constructivist Approach. *Journal of Consciousness Studies*, Vol.5 No.5–6.

C.Trevarthen, (1993) The self born in intersubjectivity: The psychology of an infant communicating. *The perceived Self ed. U.Neisser, Cambridge Univ. Press.*

C.Trevarthen, (1977) Descriptive Analyses of Infant Communicative Behavior. *Studies in Mother-Infant Interaction, ed. H.R.Schaffer, London:Academic Press.*

R.Uno and T.Ikegami (2002). Joint attention, prediction and language *the proceeding of Cognitive Linguistics, Hituji Press*