***Journal of Intelligent Learning Systems and Applications*, 2010, 2, 167-178**
doi:10.4236/jilsa.2010.24020 Published Online November 2010 (http://www.scirp.org/journal/jilsa)

Autonomous Adaptive Agent with Intrinsic Motivation for Sustainable HAI*

Takayuki Nozawa[1,#], Toshiyuki Kondo[2]

[1]Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan; [2]Institute of Symbiotic Science and Technology, Tokyo University of Agriculture and Technology, Tokyo, Japan; [#] corresponding author.
Email: nozawa@idac.tohoku.ac.jp, t_kondo@cc.tuat.ac.jp

Received March 25th, 2010; revised September 8th, 2010; accepted September 13th, 2010.

ABSTRACT

For most applications of human-agent interaction (HAI) research, maintaining the user's interest and continuation of interaction are the issues of primary importance. To achieve sustainable HAI, we proposed a new model of intrinsically motivated adaptive agent, which learns about the human partner and behaves to satisfy its intrinsic motivation. Simulation of interaction with several types of other agents demonstrated how the model seeks new relationships with the partner and avoids situations which are not learnable. To investigate effectiveness of the model, we conducted a comparative HAI experiment with a simple interaction setting. The results showed that the model was effective in inducing subjective impressions of higher enjoyability, charm, and sustainability. Information theoretic analysis of the interaction suggested that a balanced information transfer between the agent and human partner would be important. The participants' brain activity measured by functional near-infrared spectroscopy (fNIRS) indicated higher variability of activity at the dorsolateral prefrontal cortex during the interaction with the proposed agent. These results suggest that the intrinsically motivated adaptive agent successfully maintained the participants' interest, by affecting their attention level.

Keywords: *Human-Agent Interaction (HAI), Intrinsic Motivation, Reinforcement Learning, Functional Near-Infrared Spectroscopy (Fnirs)*

1. Introduction

Research on human-agent interaction (HAI) and human-robot interaction (HRI) has recently been growing and producing a wide range of applications, such as entertainment, therapeutic use, media of communication, and other kinds of assistance for intellectual activities [1]. For most of these applications, maintaining the user's interest and continuation of interaction are the issues of primary importance.

Among the various factors which affect the impression of HAI, Nakata *et al.* focused on the predictability of the behavior of agents. They experimentally studied how different degrees of randomness in the behavior affect the impression about the agents, and showed that maximum human interest is achieved by interaction with the agent of intermediate informational transmission efficiency [2,3]. Similarly, Kondo *et al.* investigated the relationship between the predictability and sustainability of the interaction, and showed that moderate degree of predictability can contribute to the sustainability [4].

However, humans get bored even with agents of moderate predictability, once they fix their mental model about the agents as such. To achieve more sustainable HAI, it will be useful to take notice of our own motivation in HAI, as well as in interaction with other human or animal. We are generally willing to continue interaction with others when it satisfies our intrinsic motivation for curiosity, exploration, manipulation, achievement, etc. [5]. Therefore, one promising approach to more "natural" and sustainable HAI would be to endow the agent with the intrinsic motivation.

Intrinsic motivation has recently been utilized in developmental robotics (also known as epigenetic robotics or ontogenetic robotics) to learn progressively from simpler to more complex situations, avoiding situations in

*This study was in part supported by "Symbiotic Information Technology Research Project" of Tokyo University of Agriculture and Technology, and also by the Grant-in-Aid for "Scientific Research on Priority Areas (Area No. 454)" from the Japanese Ministry of Education, Culture, Sports, Science and Technology. A preliminary report of this study was presented at the 16th International Conference on Neural Information Processing (ICONIP2009), Thailand, in December 2009.

which nothing can be learned [6,7]. However, effectiveness of intrinsically motivated agent on HAI is not certain because, unlike the usual cases in developmental robotics, the environment (human partner) can be highly dynamic. Indeed, the effectiveness of intrinsically motivated agent for sustainable HAI is yet to be explored.

In this study, therefore, we proposed a new model of adaptive interaction agent which learns about the human partner and behaves to satisfy its intrinsic motivation. The model's dynamical properties were analyzed by simulating interaction with several types of other agents. To investigate the model's effectiveness for enjoyable and sustainable HAI, we implemented the agent for a simple interaction setting and conducted a comparative experiment. In addition to investigating subjective impressions of the agents and the interaction log, we measured the activities of the prefrontal brain region of the participants during interaction by functional near-infrared spectroscopy (fNIRS), to study the effect of different types of agents on their cognitive states.

The rest of the paper is organized as follows. The model of intrinsically motivated adaptive agent is defined in Section 2. In Section 3, its dynamical properties are described by simulation. Section 4 explains the setting of HAI experiment. Section 5 gives the experimental results. Finally, Section 6 concludes the paper.

2. Model of Intrinsically Motivated Adaptive Agent

2.1. Adaptivity and Reinforcement Learning

We focus on discrete, turn-taking type of interaction, which means that an interaction is described from the agent's viewpoint as a sequence

$$\cdots \to s_t \to a_t \to s_{t+1} \to a_{t+1} \to \cdots, \tag{1}$$

where $s_t \in \mathcal{S}$ denotes *sensory input* from the human partner to the agent at time t, and $a_t \in \mathcal{A}$ denotes *action* of the agent at t.

An agent driven by certain motivation system—whether intrinsic or extrinsic—must be able to adapt to the environment (partner), to satisfy its motivation. We use the TD learning [8], which is a standard method in reinforcement learning (RL), to model the adaptive agent.

In the framework of RL, each input s_t is accompanied by a *reward* $r_t \in R$. When a new input s_{t+1} is obtained, the agent updates *value* of the preceding input s_t based on the stored reward r_t and the current value function $V : \mathcal{S} \to R$, as

$$\begin{aligned} \delta_t &= r_t - V(s_t) + \gamma V(s_{t+1}), \\ V(s_t) &\leftarrow V(s_t) + \alpha \cdot \delta_t . \end{aligned} \tag{2}$$

Here α is learning rate parameter of the value function, and γ is discount rate of a future reward in the present value. Values of these parameters should be determined empirically, taking the nature of the problem to be learned into account. In the initial state, without any a priori knowledge, one can set $V(s) = 0$ for all $s \in \mathcal{S}$.

Based on the value function V, the agent selects an action which is likely to maximize expected values for the next moment. The method of action selection is given in Section 2.2.

2.2. Internal Model

In many simple learning problems, the reward r_t is directly associated with s_t. In the intrinsically motivated agent, however, r_t is derived from the agent's *internal model* of the environment (human partner).

The role of the internal model is to predict what will be the next input s_{t+1} given a *context* (s_t, a_t), and how it is likely that a contextual situation (s_t, a_t) itself will take place.

Extension of the following discussion for context $(s_t, s_{t-1}, \ldots, s_{t-L+1}; a_t, a_{t-1}, \ldots, a_{t-L+1})$ with longer history length L is straight forward. However, longer history requires exponentially longer period of interaction to obtain a stable internal model.

For some specific problems, one could construct parametric internal models, which can be useful to save computational resources. Here, however, we adopt more extensive approach and define the internal model as a combination of the transition probability distribution $P_T(s_{t+1} \mid s_t, a_t)$ and the context probability distribution $P_C(s_t, a_t)$.

The internal model is updated based on the interaction history (1). As mentioned earlier, human partner can be highly dynamic. Therefore, the update of the internal model should incorporate decay of the memory. When the current context (s_t, a_t) and the new input s_{t+1} are given, the transition probability distribution P_T is updated by

$$P_T(s \mid s_t, a_t) \leftarrow \begin{cases} P_T(s \mid s_t, a_t) + \rho_T \{1 - P_T(s \mid s_t, a_t)\} \\ \qquad\qquad\qquad\qquad \text{if } s = s_{t+1}, \\ P_T(s \mid s_t, a_t) - \rho_T P_T(s \mid s_t, a_t) \\ \qquad\qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{3}$$

Similarly, the context probability distribution P_C is updated by

$$P_C(s, a) \leftarrow \begin{cases} P_C(s, a) + \rho_C \{1 - P_C(s, a)\} \\ \qquad\qquad \text{if } (s, a) = (s_t, a_t), \\ P_C(s, a) - \rho_C P_C(s, a) \\ \qquad\qquad \text{otherwise.} \end{cases} \tag{4}$$

ρ_T and ρ_C are learning rates of the internal model. Appropriate values of these parameters should be determined based on the sizes of input/output sets and the dynamic property of the partner. If the rates are too large, the agent loses the memory of interaction history too quickly, and thus fails to obtain an acceptable internal model. Due to the exponential nature of the update rules, possible criteria for the upper bounds of the parameters could be $(1-\rho_T)^{|\mathcal{S}|} > 1/2$ and $(1-\rho_C)^{|\mathcal{S}\times\mathcal{A}|} > 1/2$. If the rates are too small, on the other hand, the agent cannot catch up with the dynamic change of the human partner. Therefore, the lower bounds depend on the property of the partner, and these parameters should be empirical. The important point is avoiding extremely small values and thus giving the agent some opportunities to take initiative in the interaction (generally, this does not require fine tuning).

In the initial state, when no a priori knowledge is available, $P_T(s'|s,a) = 1/|\mathcal{S}|$ for all $(s,a,s') \in \mathcal{S}\times\mathcal{A}\times\mathcal{S}$ and $P_C(s,a) = 1/|\mathcal{S}\times\mathcal{A}|$ for all $(s,a) \in \mathcal{S}\times\mathcal{A}$.

The transition probability P_T is also used to derive *action values* for the action selection, as

$$V(a_t = a \mid s_t) = \sum_{s'} V(s') P_T(s_{t+1} = s' \mid s_t, a_t = a). \tag{5}$$

We utilize the Boltzmann action selection method, so the probability of the agent selecting action $a_t = a$, given s_t, is

$$p(a_t = a \mid s_t) = \frac{e^{V(a|s_t)/T}}{\sum_b e^{V(b|s_t)/T}}, \tag{6}$$

where T is the temperature parameter, which determines the balance between the maximization of the expected value based on the current value function and the exploration for refinement of the value function.

2.3. Intrinsic Motivation for Information Transfer

In considering the proper expression of reward for the intrinsically motivated adaptive agent, we directed our attention on the *transfer entropy* [9], which is an information-theoretic measure quantifying the causal interaction between two systems, excluding the shared information due to common history. The transfer entropy can be utilized to characterize autonomous systems [10].

From the probability distribution p of triadic interaction sequence (s_t, a_t, s_{t+1}), transfer entropy from the agent (A) to the human partner (H) is given as

$$\begin{aligned} TE_{A\to H} &= MI(s_{t+1}; a_t \mid s_t) \\ &= H(s_{t+1} \mid s_t) - H(s_{t+1} \mid s_t, a_t) \\ &= \sum_{s_t}\sum_{a_t}\sum_{s_{t+1}} p(s_{t+1} \mid s_t, a_t) p(s_t, a_t) \log \frac{p(s_{t+1} \mid s_t, a_t)}{p(s_{t+1} \mid s_t)}. \end{aligned} \tag{7}$$

Here $H(s_{t+1} \mid s_t) = -\sum_{s_t}\sum_{s_{t+1}} p(s_t, s_{t+1}) \log p(s_{t+1} \mid s_t)$ and $H(s_{t+1} \mid s_t, a_t) = -\sum_{s_t}\sum_{a_t}\sum_{s_{t+1}} p(s_t, a_t, s_{t+1}) \log p(s_{t+1} \mid s_t, a_t)$ are the conditional entropy, and $MI(s_{t+1}; a_t \mid s_t)$ is the conditional mutual information. The more the agent's action a_t has influence on the human's response s_{t+1} given the same s_t, the larger $TE_{A\to H}$ becomes (in other words, the more information the agent can transfer to the human partner).

Our intrinsically motivated adaptive agent tries to maximize this measure of influence $TE_{A\to H}$ on the human partner. This leads the following reward function

$$\begin{aligned} r_{t+1} &= \log \frac{P_T(s_{t+1} \mid s_t, a_t)}{P_{C,T}(s_{t+1} \mid s_t)} \\ &= \log \frac{P_T(s_{t+1} \mid s_t, a_t) \sum_b P_C(s_t, b)}{\sum_b P_T(s_{t+1} \mid s_t, b) P_C(s_t, b)}. \end{aligned} \tag{8}$$

Note that the reward is expressed in terms of the internal model (P_T, P_C). Maximization of the reward leads to the maximization of $TE_{A\to H}$ (note the correspondence between the reward term and the term in Equation (7), given that the internal model is sufficiently precise).

By replacing all the probability terms in Equation (7) with those of the internal model, one can also obtain the agent's *subjective transfer entropy*

$$\begin{aligned} STE_{A\to H} &= \sum_{s_t}\sum_{a_t}\sum_{s_{t+1}} P_T(s_{t+1} \mid s_t, a_t) P_C(s_t, a_t) \\ &\quad \times \log \frac{P_T(s_{t+1} \mid s_t, a_t) \sum_b P_C(s_t, b)}{\sum_b P_T(s_{t+1} \mid s_t, b) P_C(s_t, b)}. \end{aligned} \tag{9}$$

As the agent cannot directly access the probability distribution p in Equation (7), the subjective transfer entropy provides a dynamic estimate, from the viewpoint of the agent, of how much it is controlling the environment (partner). When the internal model is well adapted, the subjective transfer entropy gives a good estimate of $TE_{A\to H}$.

Let us describe through several possible situations how the reward (8) controls the behavior of the agent. First, when the agent finds an action a_t which can effectively induce otherwise rare response s_{t+1} given s_t, $P_T(s_{t+1} \mid s_t, a_t) >> P_{C,T}(s_{t+1} \mid s_t)$ and the reward is high. However, when the agent repeats the same pattern of interaction sequence (s_t, a_t, s_{t+1}) for its high value, both the numerator and denominator terms of (8) come close to 1 due to the update rules (3) and (4), so the reward and the value decrease, which correspond to the *loss of interest*, making the agent stop the repetition. On the other hand, when a_t is followed by an unexpected s_{t+1} given s_t, that is, $P_T(s_{t+1} \mid s_t, a_t) < P_{C,T}(s_{t+1} \mid s_t)$, the reward becomes negative, and such cases occur more often in the situations where the agent cannot influence the human response. Taking these considerations together, one can expected that the reward makes the agent pursue intermediate level of novelty, avoiding situations in which nothing can be learned, in a similar way as the intelligent adaptive curiosity proposed by Oudeyer *et al.* [7].

2.4. Algorithm

Combining the components described above, the operation of the intrinsically motivated adaptive agent can be described in the following procedural form:

- Initialize (s_0, a_0, r_0), the value function V, and the internal model (P_T, P_C);
- Starting from $t = 0$, repeat
  - ○ With the given context (s_t, a_t), obtain new input s_{t+1} from the environment (partner);
  - ○ Update the value function $V(s_t)$ by the TD learning method (2);
  - ○ Update the internal model (P_T, P_C) by the rules (3) and (4);
  - ○ Evaluate the reward r_{t+1} by (8), and store it for the update of value function (2) in the next time step;
  - ○ Select the action a_{t+1} using the rules (5) and (6);
  - ○ $t \leftarrow t+1$;

3. Simulation

In this section, we describe the behavior of the intrinsically motivated adaptive agent by simulating its interaction with several types of other agents.

In the following, both the intrinsically motivated agent and the other agent accept three types of inputs and take three types of actions; that is, $\mathcal{S} = \mathcal{A} = \{1,2,3\}$ for both agents. We used the parameter values shown in **Table 1** for the intrinsically motivated agent, unless stated otherwise. We utilize the transition of the action probabilities given by Equation (6) and the subjective transfer entropy given by Equation (9) to characterize the interactions.

3.1. Interaction with a Random Action Agent

Figure 1 shows the transition of action probabilities and the subjective transfer entropy of the intrinsically motivated agent interacting with an agent which selects its action a_t randomly with equal probability 1/3, regardless of the input s_t. In this case, the intrinsically motivated agent cannot control the other agent, so the action probabilities fluctuate around the uniform value of 1/3, and the subjective transfer entropy is nearly zero.

Table 1. Parameter values used for the adaptive agents in Section 3 and 4.

Parameter	Value
Learning rate of value function α	0.1
Discount rate of future reward γ	0.3
Learning rate of transition probability ρ_T	0.1
Learning rate of context probability ρ_C	0.1
Temperature for action selection T	1/30

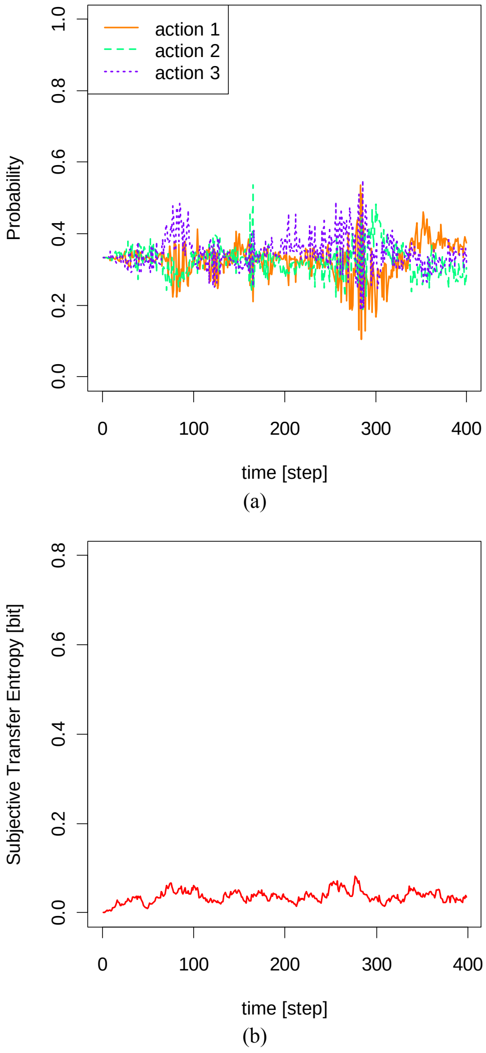


Figure 1. Transition of the action probabilities (a) and the subjective transfer entropy (b) of the intrinsically motivated adaptive agent, which is interacting with the random action agent.

3.2. Interaction with a Partially Regular Agent

Next, we study the interaction with a partially regular agent, which chooses its response a_t to an input s_t by the following response probability matrix

$$p(a_t = j \mid s_t = i) = (p_{i,j}) = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 2/3 & 0 \\ 1/3 & 0 & 2/3 \end{pmatrix}. \quad (10)$$

Note that the inputs $s_t = 2$ and 3 to this agent elicit with higher probability the responses $a_t = 2$ and 3, respectively, while $s_t = 1$ exerts no control on the agent.

Figure 2 shows the transition of action probabilities and the subjective transfer entropy of the intrinsically motivated agent interacting with the agent defined by Equation (10). The intrinsically motivated agent learns to avoid the action 1 (**Figure 2(a)**) and to achieve higher degree of control on the partner (**Figure 2(b)**). This result shows that the agent is actually capable of avoiding situations where nothing can be learned. **Figure 2(a)** also shows that the intrinsically motivated agent keeps trying to find a new controllable relationship by altering its action between 2 and 3, rather than adhering to one control strategy.

3.3. Interaction with a Fixed-Reward Adaptive Agent

Here, the intrinsically motivated agent (agent 1) interacts with another adaptive agent (agent 2), which has the same rules for the update of value function (2) and of the internal model (3), (4), and uses the same method (5), (6) for action selection, with the same parameter values in **Table 1**, but its reward r_t is directly associated with the input s_t by

$$r_t = \begin{cases} 1 & \text{if } s_t = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

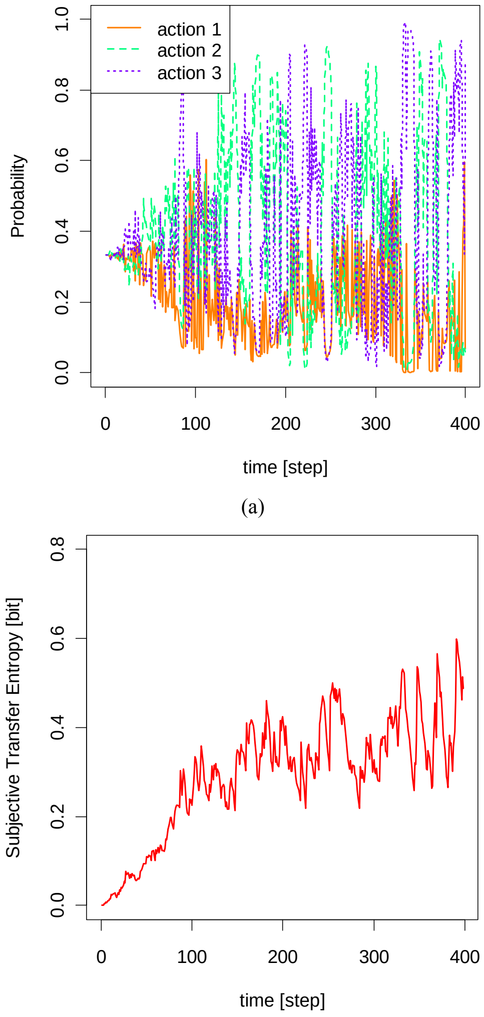
Figure 3 shows the transition of action probabilities and the subjective transfer entropy of both agents. Similar to in the interaction with the partially regular agent, the intrinsically motivated agent achieves control on the partner (**Figure 3(c)**) by altering its action strategy (**Figure 3(a)**) and thus affecting that of the partner (**Figure 3(b)**). The fixed-reward agent, on the other hand, does not exert much influence on the intrinsically motivated agent.

3.4. Interaction between Two Intrinsically Motivated Agents

Finally, we show the interaction of two intrinsically motivated agents. **Figure 4** shows the transition of action probabilities and the subjective transfer entropy of the two interacting intrinsically motivated agents. In this case, the agents competes with each other for control and keeps changing their strategies (**Figure 4(a), (b)**), so the subjective transfer entropy does not reach the level achieved in the interactions with more static agents (cf. **Figure 2(b)** and **Figure 3(c)**).

As discussed in Section 2.2, the time scale in which the intrinsically motivated agent changes its strategy depends on the learning rates of the internal model; therefore, the agent becomes slower to lose its interest in the established relationship, by decreasing the ratio of the learning rate of the context probability ρ_C to that of the transition probability ρ_T. **Figure 5** shows the transition of action probabilities and the subjective transfer entropy of the two interacting intrinsically motivated agents, whose ρ_C s were set to 0.01. This result indicates that the learning rates control the time scale of the agent's dynamics, with decreased values inducing slower transition of both the action probability distribution and the subjective transfer entropy.

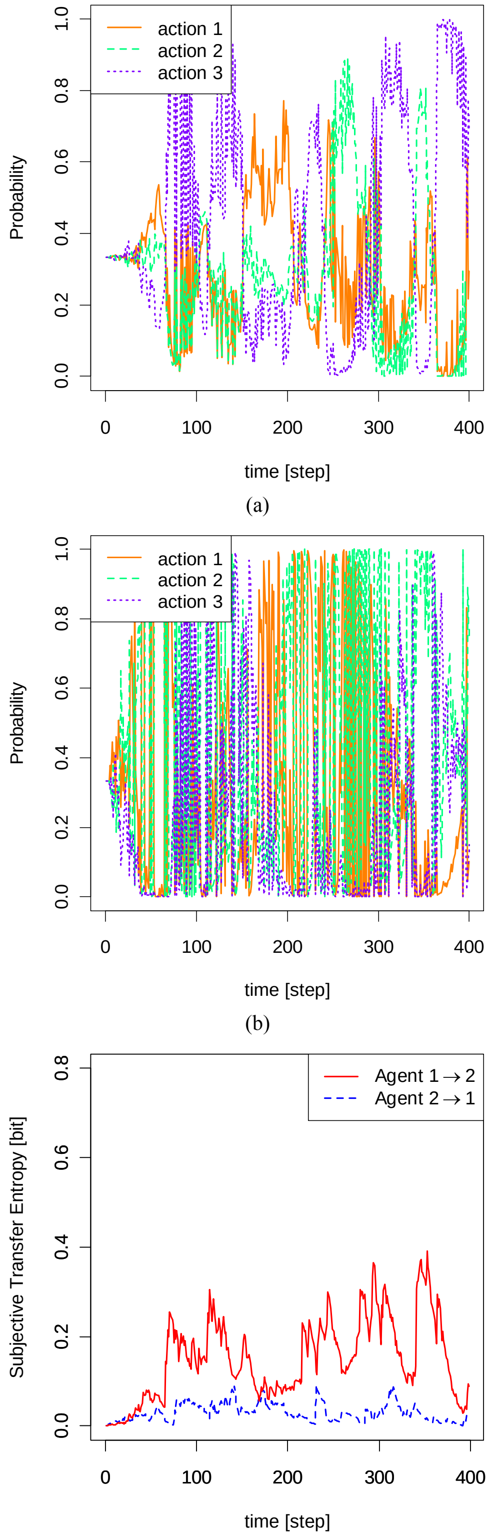
In summary, the intrinsically motivated agent demonstrated its capabilities to pursue learnable and controllable situations, and to avoid fixed relationship with the partner by altering its action strategy. These features are expected to induce the impression of sometimes unpre-



(a)

(b)

Figure 2. Transition of the action probabilities (a) and the subjective transfer entropy (b) of the intrinsically motivated adaptive agent in interaction with the partially regular agent.

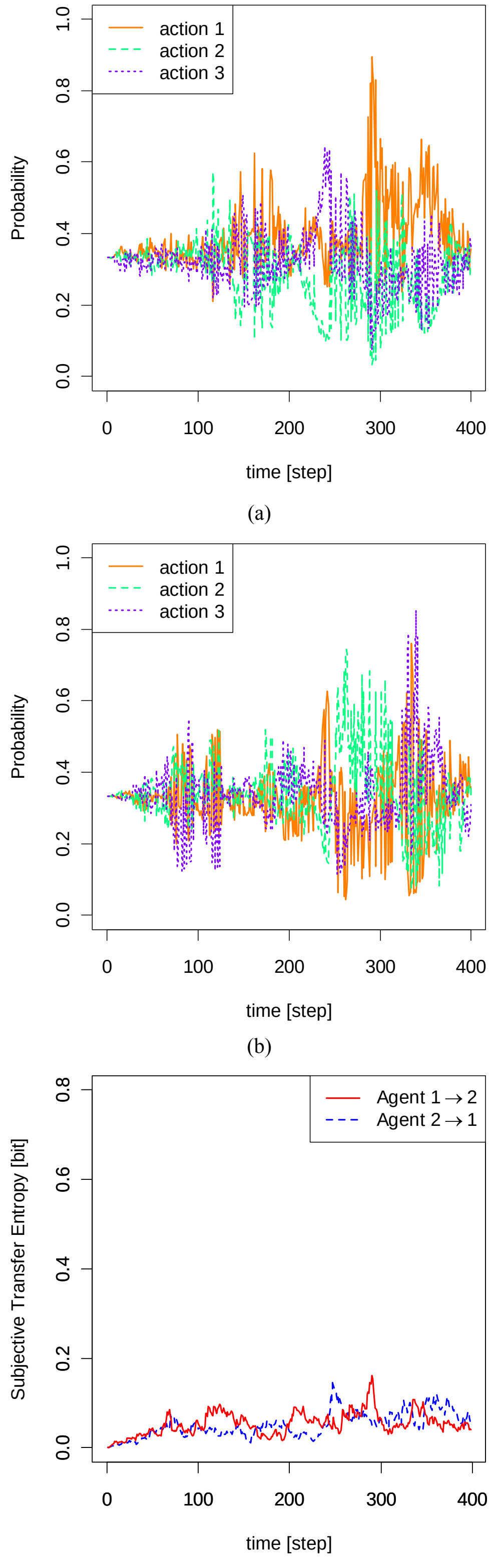


(a)

(b)

(c)

Figure 3. Transition of the action probabilities of the intrinsically motivated agent (agent 1; a), that of the fixed-reward adaptive agent (agent 2; b), and the subjective transfer entropy (c) of both agents.



(a)

(b)

(c)

Figure 4. Transition of the action probabilities of the two intrinsically motivated agents ((a) agent 1, (b) agent 2), and their subjective transfer entropy (c).

dictable yet coherent and understandable behavior, and thus would be effective in achieving more natural and sustainable HAI.

4. Experiment

To assess the effectiveness of our model of intrinsically motivated adaptive agent in HAI, we conducted a comparative experiment of three types of agents in a simple interaction design.

4.1. Interaction Design

We used a virtual agent, rather than a real robot agent, to prevent physically induced artifacts on the fNIRS measurement by minimizing the participants' movements and changes in posture during interaction with the agent [11].

A CG image of AIBO, Sony's four-legged robot, was presented on a 14.1 inch LCD display that was placed 70 cm in front of the participant sitting on a chair. Using a computer mouse, the participant clicked or dragged on the agent image. Based on the mouse-button pressing time, the agent distinguished each mouse input either as a click or as a drag (thus $\mathcal{S} = \{click, drag\}$), with the threshold of 350 ms.

The agent, in return, showed one of four actions (movies) $\mathcal{A} = \{M_1, M_2, M_3, M_4\}$; Action M_1 was to move the agent's head upward, then downward, and upward back to the neutral position. M_2 was to move its head upward and then shaking it left and right (once each side), then back to the neutral position. M_3 was to move its head up-right, back to neutral, up-left, back to neutral again. M_4 was to move its head downward, wiggle it forward and backward three times in quick succession, then move back to neutral. The meanings of the actions were left to the interpretation of each participant. Each action took 1.5 s, and the agent did not accept new mouse input till the period ends.

4.2. Agents in Comparison

The following three types of agents were compared in the interaction experiment:

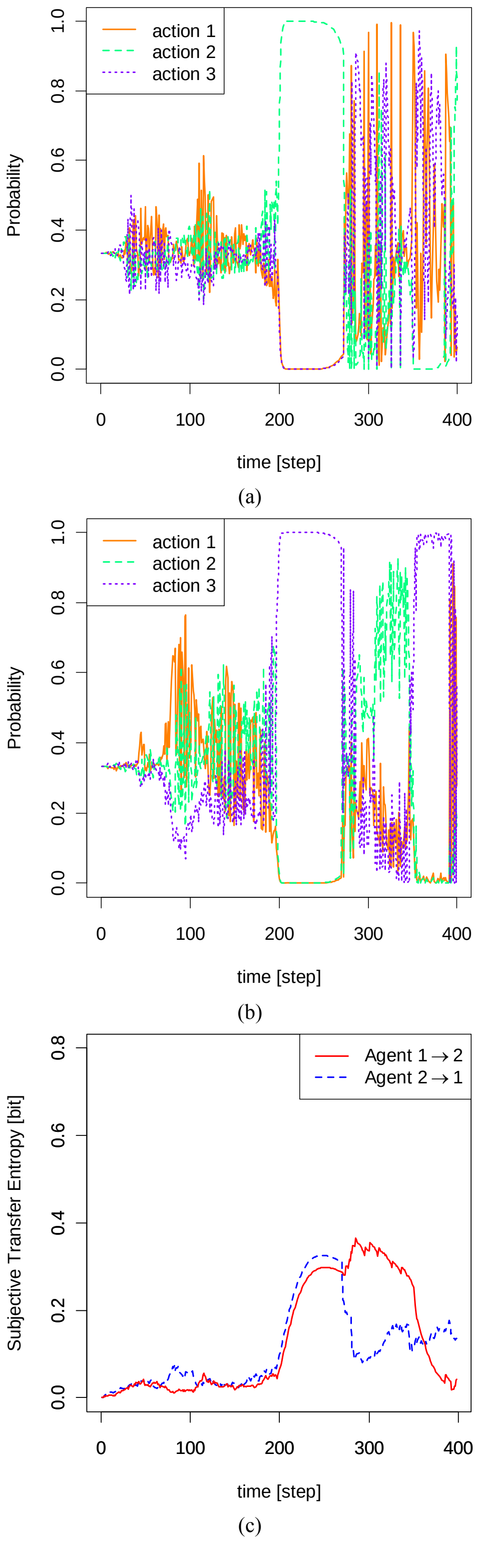
Type I was the intrinsically motivated adaptive agent defined in Section 2.

Type F was an adaptive agent which, like the fixed reward adaptive agent used in Section 3.3, had the same rules (2-6) for learning, but the reward was extrinsically given by either of the fixed functions

$$r_t^{F1} = \begin{cases} 1 & \text{if } s_t = \text{click}, \\ 0 & \text{if } s_t = \text{drag}, \end{cases}$$

$$\text{or} \qquad (12)$$

$$r_t^{F2} = \begin{cases} 0 & \text{if } s_t = \text{click}, \\ 1 & \text{if } s_t = \text{drag}. \end{cases}$$



(a)

(b)

(c)

Figure 5. Transition of the action probabilities of the two intrinsically motivated agents ((a) agent 1, (b) agent 2), and their subjective transfer entropy (c). The learning rate of the context probability $\rho_C = 0.01$ for both agents.

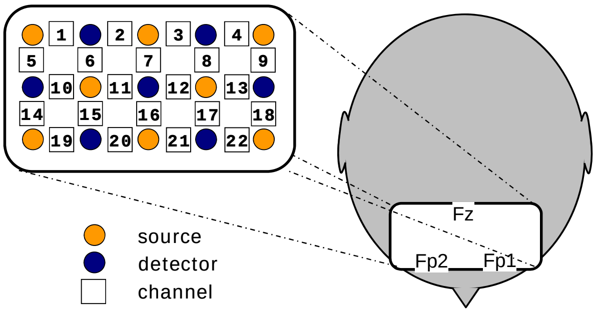


Figure 6. The arrangement of fNIRS optodes on the scalp.

Type R was an agent which selected its action a_t randomly with equal probability 1/4, regardless of the input s_t.

The parameter values in **Table 1** were used for the adaptive agents of type I and F. These parameter values were determined based on some simulations and a preliminary experiment.

4.3. Participants and Procedures

Twenty four healthy graduate or undergraduate students (23 males and 1 female, all right handed, mean age 21.7 ± S.D. 1.7 years) participated in the experiment. All participants were explained about the experiment before giving written informed consent. This study was approved by the ethics committee of the Tokyo University of Agriculture and Technology.

Before the experiment, the participants were familiarized with the operation, by a 5 min practice session with an agent that showed all the four actions in an order when the mouse was clicked, and in the reverse order when dragged.

The participants were instructed to freely set and change their aims of interaction. Each participant had interaction sessions with all the three types of agents. They were divided into four groups (six participants each), by the two orderings of agent types, (R, I, F) or (F, I, R), and by the two kinds of reward r^{F1} or r^{F2} in (12) for type F agent.

Each interaction session consisted of 1 min fixation phase, 15 min interaction phase, and again 1 min fixation phase. During the fixation phases, the participants were instructed to fixate their attention on a cross shown at the center of the display. Each interaction session was followed by 10 min rest period, during which they were asked to complete the questionnaire.

4.4. Questionnaire

After each session, the participants were asked to describe their impression about the agent they interacted with. After the second and third sessions, they were also asked to answer a Likert scale questionnaire comparing the last two agents they interacted with. The questionnaire consisted of 16 items with eight viewpoints, each item with seven rating levels from "strongly disagree" (−3) to "strongly agree" (+3). The eight viewpoints were:

1. enjoyable,
2. charming,
3. lively,
4. soothing,
5. consistent,
6. obedient,
7. insightful to your intention, and
8. desirable for longer period of interaction.

For each of the eight viewpoints (adjectives), two items —"The latter felt *more* {adjective} than the former." and "The latter felt *less* {adjective} than the former." —were presented. This was to balance the influence of positive/negative expressions, and to check the consistency of each participant's ratings. The items were arranged in a randomized order.

4.5. Interaction Log

In regard to the actions of participants, timing and types of mouse operations were recorded. For the agents, timing and types of movie actions were recorded. For the adaptive agent of type I and F, the sequence of rewards r_t, the estimated value function V, and the internal model (P_T, P_C) were also logged.

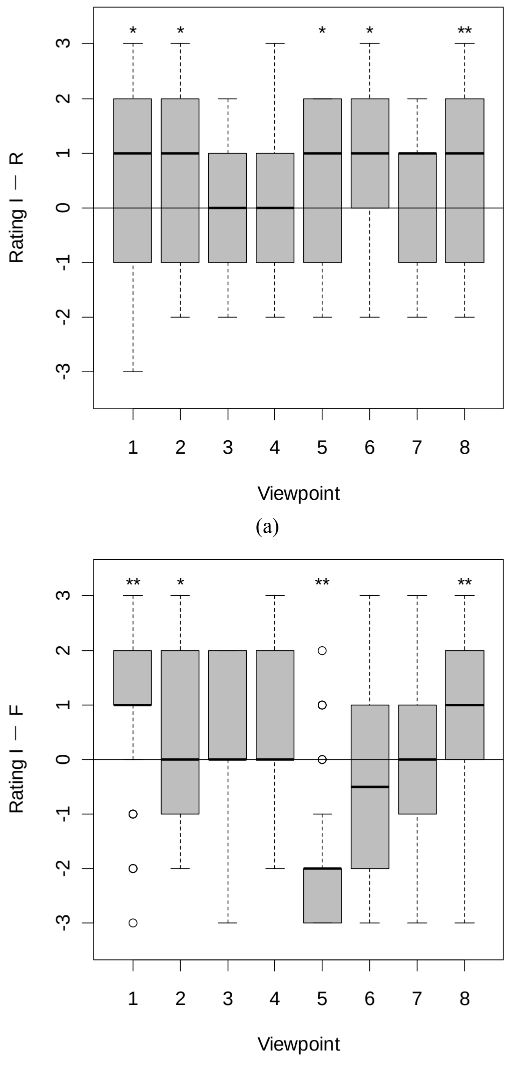
From these data, we examined statistics of human/agent actions, information theoretic measures of interaction, such as mutual information, distinguish ability, controllability (dyadic) [3] and transfer entropy (triadic), and dynamics of these measures.

4.6. fNIRS Measurement

Prefrontal region of human brain plays significant roles for attention control, working memory, executive function, etc. [12], which will be important for sustainable HAI. Therefore, we measured the activity in prefrontal brain region of twelve participants during the interaction sessions by functional near-infrared spectroscopy (fNIRS).

Like functional magnetic resonance imaging (fMRI), fNIRS assesses brain activities based on hemodynamic responses. It enables us to measure the changes in concentration of oxygenated, deoxygenated, and total hemoglobin (oxy-Hb, deoxy-Hb, and total-Hb) within cortical tissue. In the analysis, we focused on the oxy-Hb, as it is suggested to be the most sensitive indicator of activity-dependent hemodynamic changes [13].

We used a multi-channel fNIRS instrument (FOIRE-3000, Shimadzu Co., Japan). Eight source and



(a)

(b)

Figure 7. Boxplot of the ratings on comparative impressions with respect to eight viewpoints. (a) the agent type I vs. R. (b) the agent type I vs. F. See Section 4.4 for the contents of the viewpoints.

seven detector optodes were placed on the prefrontal regions, covering Fp1, Fp2 and Fz positions of the international 10-20 system, with a total of 22 channels (**Figure 6**). The data were acquired at a sampling period of 70 ms. To reduce instrumental and physiological noise, the signals were band-pass-filtered with Chebyshev type II filter of 4-th order with cut-off frequencies of 0.7 and 0.002 Hz, pass-band ripple 5 dB.

To avoid physically induced artifacts as much as possible, the participants were asked to assume a preferred posture and retain it for the entire duration of the experiment. To avoid the displacement of optodes, the participants kept the fNIRS optodes on throughout all the three interaction sessions.

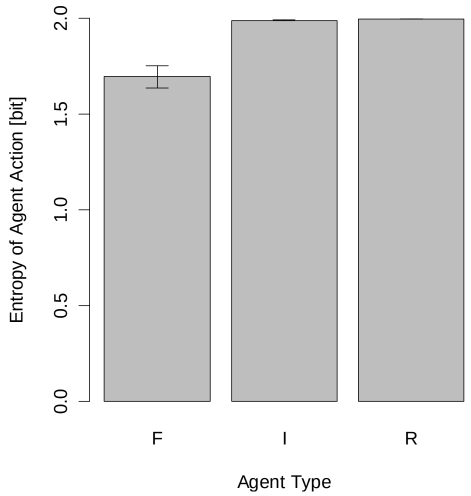


Figure 8. Entropy of agent actions, which was calculated for each interaction session and then averaged for each agent type over participants. Error bars indicate the standard error of the mean.

5. Results

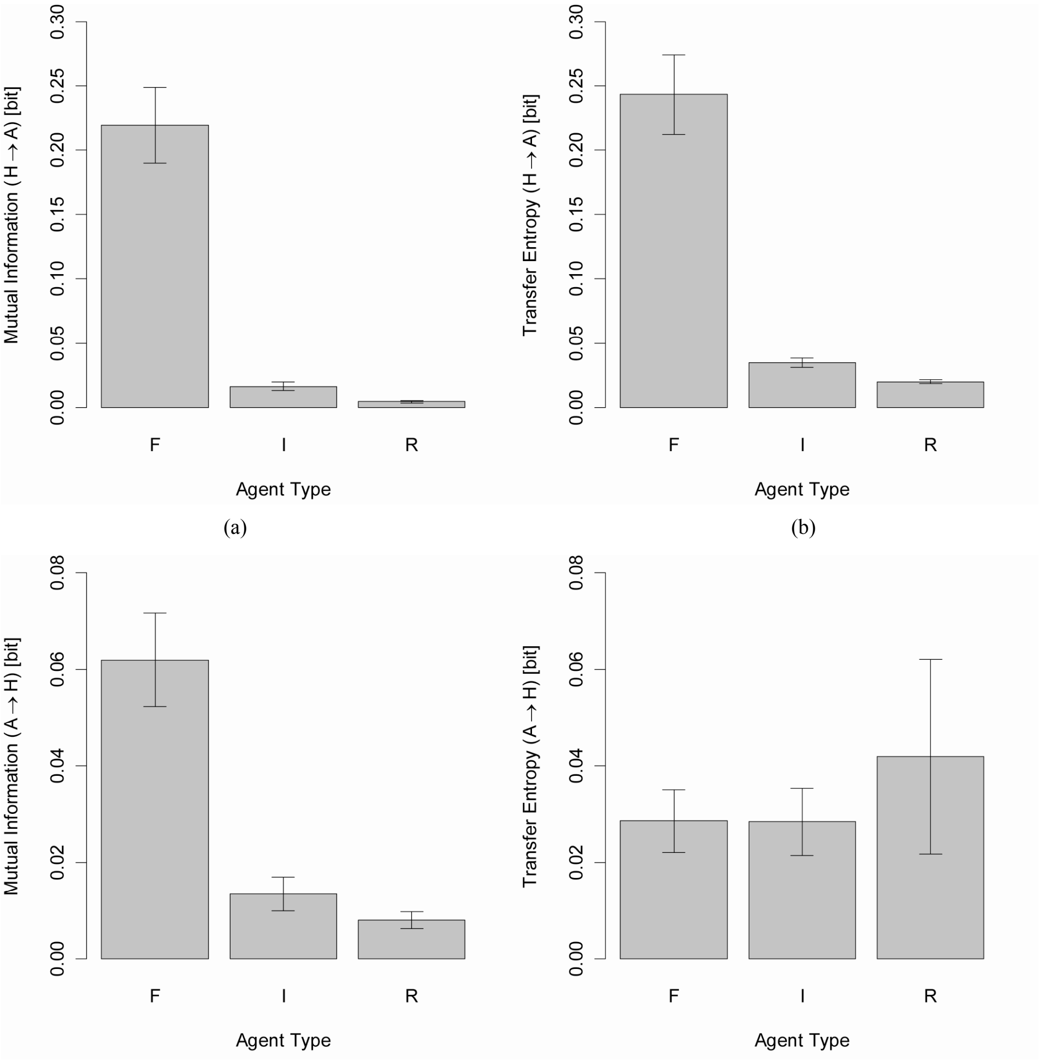
For three participants (one with NIRS measurement), over 95 percent of their mouse actions were either click or drag in at least one session. Therefore, they were judged to have conducted the interaction improperly, and excluded from the following analyses.

5.1. Subjective Impressions

Figure 7 shows the distribution of the comparative ratings between type I and the other two types, with respect to the eight viewpoints given above. The viewpoints with which Wilcoxon signed rank test (null hypothesis: the rating is symmetric about 0) indicated significant difference with level $p < 0.05$ is marked with "*", and those with level $p < 0.01$ are marked with "**". This result shows that the intrinsically motivated adaptive agent gave impressions of higher enjoy ability, charm, and sustainability than the other two types of agents (viewpoint 1: $p = 0.023$ for type I vs. R and $p = 0.002$ for type I vs. F; viewpoint 2: $p = 0.032$ for type I vs. R and $p = 0.030$ for type I vs. F; viewpoint 8: $p = 0.006$ for type I vs. R and $p = 0.001$ for type I vs. F).

5.2. Statistics of Actions

The average number of interactions in a session was 383.9. There were no significant differences in the number among the three agent types. Regarding human actions, the average number of click was 256.9, and that of drag was 127.0. The ratio of click to drag did not show significant differences among the agent types, either. **Figure 8** shows the difference of average entropy of agent actions for the agent types. The frequency



(a)

(b)

(c)

(d)

Figure 9. Static estimation of (a) mutual information $MI(s_t, a_t)$, (b) transfer entropy $TE_{H \to A}$, (c) mutual information $MI(a_t, s_{t+1})$, and (d) transfer entropy $TE_{A \to H}$. Error bars indicate the standard error of the mean.

distributions of the agent actions were more biased in the inter actions with type F agent ($p < 0.001$ for F vs. I and F vs. R, $p = 0.021$ for type I vs. R).

5.3. Characteristics of Information Transfer

To compare the features of interaction with the different types of agents from the information theoretic viewpoint, first, we computed the frequency distributions of dyadic pairs (s_t, a_t) , (a_t, s_{t+1}) and triadic interactions (a_{t-1}, s_t, a_t), (s_t, a_t, s_{t+1}) over each interaction session. Using these, we calculated mutual information and transfer entropy and compared them among the agent types (**Figure 9**; results of the distinguish ability and controllability were omitted because they were qualitatively equivalent to those of the mutual information).

Figure 9(a) and **(b)** show that type I agent caused intermediate level of information transfer from the participant to the agent, meaning that the predictability of its actions for human was also intermediate between the other two types of agents. Multiple pair wise comparisons (Wilcoxon signed rank test with Holm's multiple test correction) showed significant differences in $MI(s_t, a_t)$ and in $TE_{H \to A}$ between the agent types ($p < 0.001$ for all pairs).

For the information transfer from agent to human, the mutual information and transfer entropy exhibited different characteristics (**Figure 9(c)** and **(d)**). Multiple pair wise comparisons showed significantly larger $MI(a_t, s_{t+1})$ with type F than with the other two types ($p < 0.001$ for F vs. I and F vs. R, $p = 0.243$ for I vs. R),

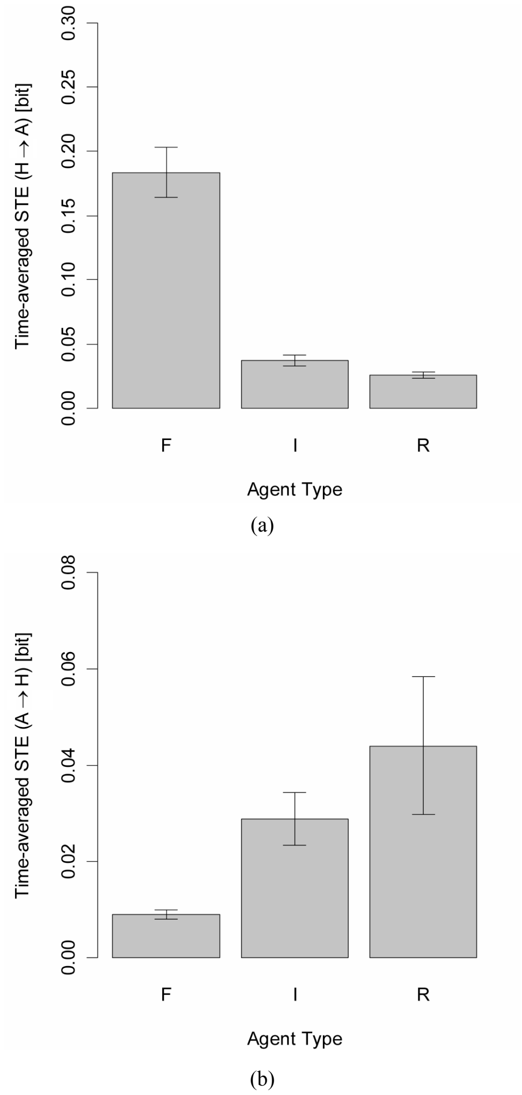


Figure 10. Time-averaged subjective transfer entropy from the participants to the agents (a), and from the agents to the participants (b). Error bars indicate the standard error of the mean.

but no significant differences were found in $TE_{A \to H}$ between any pairs of agent types. We also note that $TE_{A \to H}$ was significantly larger than $MI(a_t, s_{t+1})$ for type I agent (Wilcoxon signed rank test, $p < 0.001$).

To capture differences in dynamic aspects of the interaction more accurately, we evaluated the transition of subjective transfer entropy for all sessions. In addition to type I and F agents, the internal models updated by the rules (3) and (4) with the values of the learning rates (ρ_T, ρ_C) in **Table 1** were also hypothesized in type R agent and in the participants as well, and they were used to calculate the subjective transfer entropy by Equation (9).

Figure 10 shows the differences of the time-averaged $STE_{H \to A}$ and $STE_{A \to H}$ for the three types of agents. $STE_{H \to A}$ in **Figure 10(a)** showed a similar result to **Figure 9(b)**. $STE_{A \to H}$ in **Figure 10(b)**, on the other hand, manifested significant differences between all agent types ($p < 0.001$ for F vs. I and F vs. R, $p = 0.003$ for I vs. R). This indicates that once s_t was given, the action a_t of type I agent had more influence on the human response s_{t+1} than that of type F agent. A possible interpretation for the highest $STE_{A \to H}$ in the interaction with type R agent is that as the participants could not find any strategy in the agent, they invented an imaginary relationship and played their subservient roles.

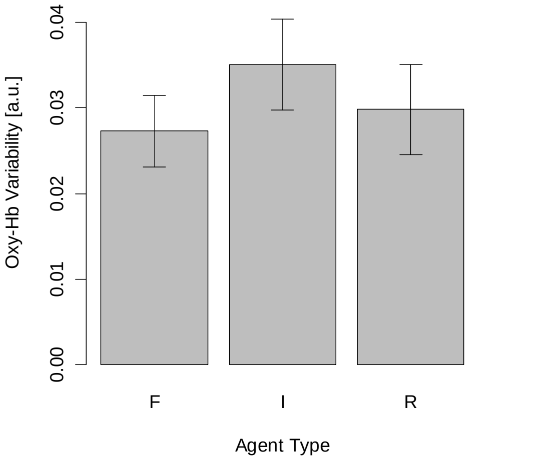


Figure 11. Variability of fNIRS oxy-Hb signal from a lower left channel. Error bars indicate the standard error of the mean.

These results suggest that the intrinsically motivated adaptive agent induced better subjective impressions by achieving a balanced information transfer with the human partners.

5.4. Variability of Activity in Prefrontal Cortex

We evaluated the variability of activity in the prefrontal region by the standard deviation of oxy-Hb signals from each of the 22 channels. By multiple pairwise comparisons (Wilcoxon signed rank test with Holm's multiple test correction), the variability showed significantly higher values with type I agent ($p = 0.008$ for I vs. F, $p = 0.006$ for I vs. R, $p = 0.375$ for F vs. R; see also **Figure 11**) at a lower left channel (channel 22 in **Figure 6**). The channel was overlapped with the dorsolateral prefrontal cortex (DLPFC), which is involved in controlling and sustaining attention [12]. Therefore, this higher variability suggests that the intrinsically motivated adaptive agent successfully kept affecting the participants' attention level.

6. Conclusions

To achieve sustainable HAI, we proposed a new model of intrinsically motivated adaptive agent, which tries to maximize its influence on the human partner. The simulation demonstrated how the model tries to keep satisfying its motivation by pursuing new relationships with the partner and by avoiding situations where nothing can be learned. To assess the effectiveness of the intrinsically motivated adaptive agent, we conducted a comparative HAI experiment with three types of agents. The results showed that the model was effective in inducing subjective impressions of higher enjoy ability, charm, and sustainability. Information theoretic analysis of the interaction suggested that a balanced information transfer between the agent and human partner would be important for sustainable HAI. The participants' brain activity measured by fNIRS indicated higher variability of activity at the left DLPFC during interaction with the proposed agent, suggesting that the model kept affecting the participants' attention level.

Unlike the models of intrinsically motivated learner in developmental robotics [6,7], our model did not incorporate the extension of dimensions in input-action space and the internal model space. Such a developmental aspect will be effective for longer term sustainable HAI, though experimental assessment of its effectiveness would become more qualitative.

REFERENCES

[1] T. Fong, I. Nourbakhsh and K. Dautenhahn, "A Survey of Socially Interactive Robots," *Robotics and Autonomous Systems*, Vol. 42, No. 1, 2003, pp. 143-166.

[2] T. Nakata, T.-S. Ko, T. Mori and T. Sato, "Informational Analysis on Impression of Human Robot Interaction," *Journal of Robotics Society of Japan*, Vol. 19, 2001, pp. 667-675.

[3] T. Nakata, "Information Transmission and Impression Formulation in Human-Robot Interaction," 2004. http://staff.aist.go.jp/toru-nakata/hri/HRI-Info.html

[4] T. Kondo, D. Hirakawa and T. Nozawa, "Sustainability and Predictability in a Lasting Human-Agent Interaction," *Proceedings of* 8*th International Conference on Intelligent Virtual Agents*; *Springer*, *Lecture Notes in Computer Science*, Japan, 2008, pp. 505-506.

[5] E. L. Deci and R. M. Ryan, "Intrinsic Motivation and Self-Determination in Human Behavior," Plenum, New York, 1985.

[6] A. Barto, S. Singh and N. Chentanez, "Intrinsically Motivated Learning of Hierarchical Collections of Skills," *Proceedings of* 3*rd International Conference on Developmental Learning*, UCSD Institute for Neural Computation, La Jolla, CA, 2004, pp. 112-119.

[7] P. Oudeyer, F. Kaplan and V. V. Hafner, "Intrinsic Motivation Systems for Autonomous Mental Development," *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 2, 2007, pp. 265-286.

[8] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," MIT Press, Cambridge, MA, 1998.

[9] T. Schreiber, "Measuring Information Transfer," *Physical. Review Letters*, Vol. 85, No. 2, 2000, pp. 461-464.

[10] N. Bertschinger, E. Olbrich, N. Ay and J. Jost, "Autonomy: An Information Theoretic Perspective," *BioSystems*, Vol. 91, No. 2, 2008, pp. 331-345.

[11] T. Nozawa and T. Kondo, "A Comparison of Artifact Reduction Methods for Real-Time Analysis of Fnirs Data," *Proceedings of* 13*th International Conference on Human-Computer Interaction, Lecture Notes in Computer Science*, Springer, Heidelberg, Vol. 5618, 2009, pp. 413-422.

[12] E. K. Miller and J. D. Cohen, "An Integrative Theory of Prefrontal Cortex Function," *Annual Review of Neuroscience*, Vol. 24, No. 1, 2001, pp. 167-202.

[13] Y. Hoshi, "Functional Near-Infrared Spectroscopy: Current Status and Future Prospects," *Journal of Biomedical Optics*, Vol. 12, No. 6, 2007, pp. 062106-1-9.