The hesitation of a robot: a delay in its motion increases learning efficiency and impresses humans as teachable

TANAKA Kazuaki, OZEKI Motoyuki, and OKA Natsuki
Graduate School of Science and Technology, Kyoto Institute of Technology
Email: {d8821007, ozeki, nat}(at)kit.ac.jp

***Abstract*—If robots learn new actions through human-robot interaction, it is important that the robots can utilize rewards as well as instructions to reduce humans' efforts. Additionally, "interval" which allows humans to give instructions and evaluations is also important. We hence focused on "delays in initiating actions" and changed them according to the progress of learning: long delays at early stages, and short at later stages. We compared the proposed varying delay with a constant delay by an experiment. The result demonstrated that the varying delay improves learning efficiency significantly and impresses humans as teachable.**

***Keywords*-hesitation; delay; learning efficiency; teachability;**

I. INTRODUCTION

In the future, robots will help us in our daily life. We believe that robots should learn desirable behavior through human-robot interaction. It therefore is important that the robots can utilize rewards as well as instructions to reduce humans' efforts. Additionally, "interval" which allows humans to give instructions and evaluations is also important.

We hence focused on "delays in initiating actions" and changed them according to the progress of learning: long delays at early stages, and short at later stages. In other words, if a robot is not sure about its action, it initiates the action laggardly, but if it is confident about its action, it initiates the action immediately. the delay is inspired by the hesitation of humans when they are unsure of their actions, and so we believe that the delay is reasonable in representing the hesitation of a robot.

In our previous work [1], we conducted experiments on teaching AIBO to shake hands under three conditions: Varying Condition: the delays vary in accordance with the progress of learning (0.1 to 3.1 sec), Quick Condition: the delays are short constant (0.4 sec), and Slow Condition: the delays are long constant (2.8 sec). Under Quick Condition, it was difficult for participants to give instructions at appropriate timing. Under Slow Condition, participants could give instructions but AIBO irritated them because there were long delays even after the learning progressed. Additionally, the number of interactions decreased since the delays were constantly long during the whole experiment. In contrast with these conditions, under Varying Condition, participants could give instructions at appropriate timing in early stages because the delays were long enough, and after the learning progressed, AIBO initiated actions immediately so there was no irritation. As a result, AIBO under Varying Condition was the most efficient learner and impressed participants as teachable.

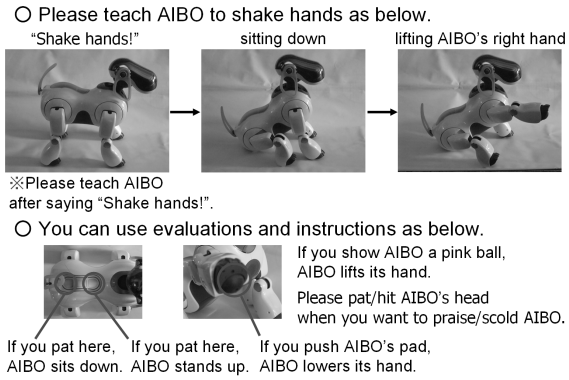

Figure 1. The procedure of shaking hands and evaluations and instructions available for the training.

In this work, we compare the Varying Condition with Constant Condition under which the delays are set at medium constant ($(0.4+2.8)/2 = 1.6$ sec), and demonstrate that the change of the delay is important for the learning efficiency and the impression.

II. EXPERIMENTAL TASK

We asked six participants to teach AIBO to shake hands using evaluations (patting/hitting AIBO's head) and instructions (e.g. patting AIBO's back, showing a pink ball, etc.) as shown in Fig. 1. AIBO selects an action from seven candidates (e.g. sitting down, lifting its hand/foot, etc.) and initiates the action after a delay specified in each condition.

III. THE METHOD OF LEARING AND DECIDING THE DELAY

A. The learning method

In this work, we employ Q-learning [2], one of the reinforcement learning algorithm. In Q-learning, the action value $Q(s, a)$, which is the value of an action a in a state s, is updated based on rewards r, and the best action in each state is found by trial and error.

$$Q(s, a_n) \leftarrow Q(s, a_n) + \alpha\{r + \gamma \max_a Q(s', a) - Q(s, a_n)\} \quad (1)$$

If participants give AIBO positive/negative reward patting/hitting, $Q(s, a)$ is updated by equation 1 using reward $r = +1.0/-1.0$. We set the learning rate α at 0.1, and discount rate γ at 0.5.

If AIBO is given an instruction I_n, AIBO immediately acts the corresponding action a_n, and after a transition to a next state s', $Q(s, a_n)$ is updated by equation 1 using reward $r = +1.0$ as if participants gave a positive reward.

Additionally, we employ Boltzmann selection, one of the method of selecting actions, and set Boltzmann temperature at 0.3. In Boltzmann selection, the selection probability $P(s, a_n)$ of the action a_n is calculated from $Q(s, a_n)$. AIBO selects an action in accordance with $P(s, a_n)$.

B. The method of deciding the delay

In this work, we compare the Varying Condition (hereafter called VC) with Constant Condition (hereafter called CC). Under VC, the delay $D(s, a_n)$ is calculated by equation 2 using the selection probability $P(s, a_n)$.

$$D(s, a_n) = d_{min} + d_{max}/(1 - e^{-cd_{max}\{0.5 - P(s,a_n)\}}) \quad (2)$$

We set the minimum delay d_{min} at 0.1 second, the maximum delay d_{max} at 3.1 second, and a constant value c at 0.4. Under CC, we set the delay at 1.6 second regardless of the selection probability.

IV. Experimental results and Discussions

A. Learning Efficiency

Fig. 2 shows the progress of the selection probabilities of correct actions (a_1: sitting down, a_2: lifting its right hand) at each state (s_1: standing, s_2: sitting) under each experimental condition (VC: Varying Condition, CC: Constant Condition). The horizontal axis represents the elapsed time of experiments. The vertical axis represents the selection probabilities of correct actions. The points plotted on the figure are averages of the results of the six participants.

Furthermore, we conducted a three-factor analysis of variance to compare the averages of learning efficiency, that is selection probability, that rose during 10 minutes. The three factors were the experimental conditions, the order of experiments (VC to CC, CC to VC), and the states (s_1, s_2). The result showed that there were significant differences in the experimental conditions ($F(1, 4) = 14.122$, $p < .05$) and in the states ($F(1, 4) = 85.577$, $p < 0.001$), and there was no significant difference in the order of experiments ($F(1, 4) = 2.152$, $n.s.$). We thus can conclude that the learning efficiency of VC is better than CC regardless of the order of experiments. The reason is that CC is just "too much for one, not enough for the other".

Under CC, participants sometimes gave evaluations/instructions at inappropriate timing because the delays were too short in early stages. Moreover, the delays became unnecessary at later stages since the number of evaluations/instructions given by participants were gradually decreased as learning progressed. In contrast, under VC, participants could give evaluations/instructions at appropriate timing because the delays were long enough in early stages, and AIBO came to initiate actions immediately as learning progressed.

B. Teachability

We asked participants to answer a questionnaire to evaluate impression on animacy, likeability, intelligence, and teachability of the robot. The items of the questionnaire except teachability were made after [3]. In this paper, we show only the result of teachability in Fig. 3. The points

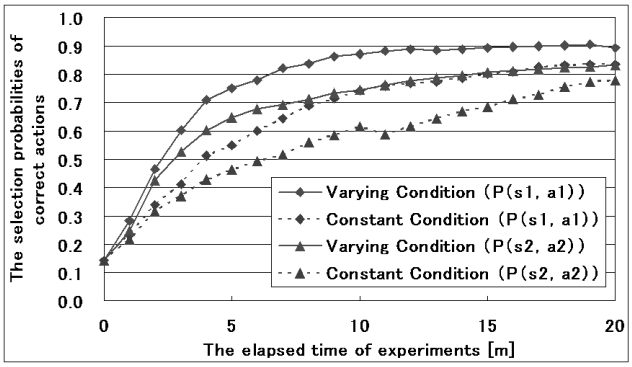

Figure 2. The comparison of the learning curves under the two conditions.

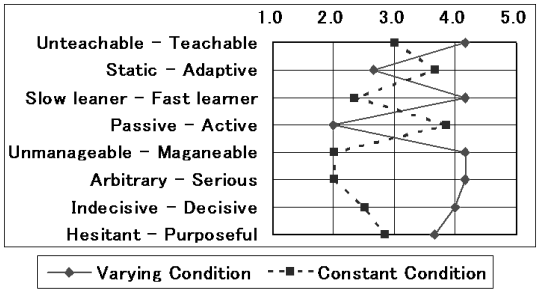

Figure 3. The impression on teachability of AIBO under each condition.

plotted on the figure are averages of the evaluations of the six participants. As shown in the figure, VC got better evaluations than CC at most items ($6/8$). We consider that the difference between the two conditions came from the same cause as one discussed in Section IV-A.

However, VC received negative evaluations with two items, static and passive. Under VC, AIBO completed the learning early and repeated the learned same actions during the rest of experiment. It is possible that the participants thus got static and passive impression on AIBO under VC.

V. Conclusion

We pointed out the importance of timing in human-robot interaction, and demonstrated experimentally that an appropriate delay in action of a robot accelerates its learning and improves its teachability.

Acknowledgment

This work was supported by JSPS KAKENHI 17500093 and 21500137.

References

[1] Tanaka, K. and Oka, N., An experimental valuation of a robot who "hesitates" in human-robot interaction, HAI2008, 2B–2, 6 pages, 2008. (in Japanese)

[2] Watkins, C. J. C. H. and Dayan, P., Q-learning, Machine Learning, Vol. 8, No. 3–4, pp. 279–292, 1992.

[3] Bartneck, C., et al., Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, International Journal of Social Robotics, Vol. 1, No. 1, pp. 71–81, 2009.