The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

An action planning model using short-term and long-term memory information during learning of sequential procedures

Nao Tomoda, Yuta Kakimoto, and Osamu Araki

Department of Applied Physics, Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan

Abstract: To make an action plan, it is thought that the brain uses memory systems. Thus, we propose an action plan model, which deals with the physiological experiments to push buttons in the correct order. In the model, there are two independent action-planning systems of long-term memory and working memory. When the stimulus set is input, they propose action plans, and the selection is decided in a competitive way via the value of estimation parameters. As a result, the model reproduces similar behaviors to the biological data. Especially our model make errors at first but it gradually learns correct responses by trial and error utilizing the memory systems as the monkey did in the physiological experiment. The results suggest that SMA and pre-SMA may have close relationships to the entries to long-term memory and working memory systems in the brain.

Keywords: action plan, long-term memory, Pre-SMA, sequential procedure, short-term memory, SMA

1. INTRODUCTION

In the physiological experiments of Hikosaka et al., 2 of 16 (4 by 4 matrix) buttons were illuminated (this matrix is called a "set") simultaneously and the monkey had to push them in a predetermined order. The monkey was required to push the buttons in 5 consecutive sets (called a "hyperset") [1]. When a hyperset which the subject has already learned is the stimulus, the task is called a "learned task". And when a hyperset consists of only novel sets, it is called a "new task". It is reported that SMA and pre-SMA are specifically activated by the case of the learned task and new task, respectively [1][3][4]. When a hyperset of which a few sets in a learned hyperset are replaced by novel sets, the task is called "modified task". Although monkey made mistakes in the early trials in the modified task, it gradually became to push the buttons correctly. Intriguingly, in the early trials, the monkey sometimes pushes unilluminated buttons in response to a novel set [1]. We think that this is because it may associate the previous stimulus sets with an action to be done next from the learned memory of a series of actions even if it actually encounters a novel stimulus.

Additionally, in the modified task, the neurons in pre-SMA become activated in response to the learned sets as well as the novel sets [1]. This implies that the activation in pre-SMA in response to the learned sets increases while the monkey is learning the modified task.

On the other hand, Nakahara et al. reproduced the behaviors and neural activities using a neural network model of the basal ganglia and the cortices in the reinforcement learning framework [2]. But the model does not attend the kinds of monkey's errors, where all of them are treated as learning errors. Thus, we pay attention to the error that is sometimes observed in the modified task, which makes the monkey to push an unilluminated buttons. From a new viewpoint to explain this error behavior, the

purpose of this study is to propose the simple functional model, which reproduces this error as well as the other behaviors observed in the previous experiments.

2. OUTLINE OF THE MODEL

2.1 Task for computer simulations

According to the previous experiments, we assume a task as follows: the subject is required to push the buttons in the correct order in 5 sets, treated as a sequential procedure, in a trial. Since 2 of 16 (4 × 4 matrix) square buttons are illuminated evenly on a set, the subject has to learn the order from the visual pattern of illuminated positions by trial and error. If it makes a mistake, the trial returns back to the first stage, but the pattern of five stimuli do not change within the same task. Thus five consecutive sets (hyperset) are presented in a fixed order. The task is repeated until the subject completed the hyperset successfully for a total of 10 trials.

2.2 Structure of the model

Fig.1 shows the outline of our proposed model. The model mainly consists of three systems as follows: short-term memory system (STMS), long-term memory system (LTMS), and decision system of strategy (DSS). STMS proposes an action plan based on the short-term memory especially dealing with stimuli in the current task, while LTMS tries to retrieve an action sequence in the long-term memory over tasks. These systems work in parallel and independently. Actually only one action plan should be selected to execute in DSS. For each stimulus set, DSS stochastically and alternatively selects one of LTMS or STMS. The probability of the selection is updated depending on the estimate of decisions and their results in response to the stimulus sets.

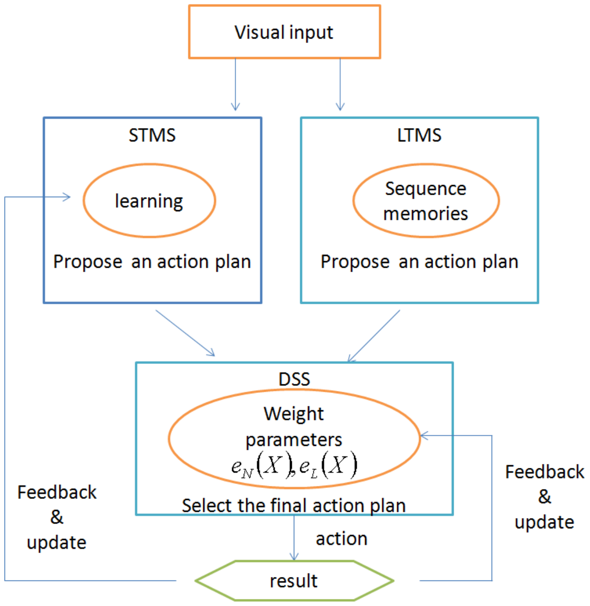


Fig.1. Outline of the model

2.3 Short-term memory system

STMS proposes one action plan from two options of action orders based on the short-term memory for the current stimulus (illuminated two buttons). In response to a novel stimulus, the probability of selection is even. As the monkey learns stimulus-response relationships, the probability to press the buttons correctly becomes higher. The probability will be influenced by how many times it has received the stimulus so far. Thus, we assume that the probability of proposal of the correct action plan for a set in STMS is updated by the short-term experience as follow

$$p_{set}(n+1) = p_{set}(n) + \gamma \left(\alpha_{MAX} - p_{set}(n)\right) \quad (1)$$

where n is the number of input of the stimulus sets, α_{MAX} is the maximum value of the probability, and γ is a learning rate.

2.4 long-term memory system

LTMS also proposes one action plan from the two options based on the long-term sequential memory. When the model recognizes that the first set is a learned stimulus, we assume that this memory system recalls 5 series of actions responsive to current and succeeded stimuli. Therefore, LTMS is completely different from the STMS in that LTMS suggests plans for supposed stimuli in the near future.

Additionally, this memory has been constructed for a long time such as a year or two years. Thus, the probability to suggest the correct order for the learned hyperset is assumed to be very high rate. Thus, we let this higher probability be P_l. When the hyperset of stimuli includes a few new sets after the first set (i.e. modified task), however, LTMS at first proposes to push an unilluminated button and the panel does not respond. For this error, LTMS is temporarily suspended and STMS is driven to deal with this stimulus.

2.5 decision system of strategy

DSS selected the final action plan to execute from the plans proposed by STMS and LTMS. We assume two weight parameters $e_S(X)$ and $e_L(X)$ to represent the reliability of STMS and LTMS, respectively. The strategy with a large value of this parameter assumes to be more reliable. If the stimulus is a learned one and the result of LTMS strategy is correct, $e_L(X+1) = e_L(X) + u_1$. But, if it was wrong, $e_L(X+1) = e_L(X) - u_2$. If LTMS is suspended in case of modified hyperset as shown above, $e_L(X+1) = e_L(X) - u_3$ .If STMS strategy is selected and the result is correct, $e_S(X+1) = e_S(X) + u_4$, otherwise $e_S(X+1) = e_S(X) - u_5$.

Initially, $e_S(0) = e_L(0) = 0$. u_1, u_2, u_3, u_4 ,and u_5 are positive constant parameters. According to the physiological experiment [1][3], the monkey seemed to infer whether the current task is a new or learned one when it watched the first stimulus set. We think that this influence strongly biases the monkey's behavior to coming stimuli. Thus, we assume that $e_L(1) = e_L(0) - b$ if the first set is new, and $e_L(1) = e_L(0) + b$ if it is already learned.

The larger the value of $e_S(X)$ or $e_L(X)$ is, the more the probability of selection of corresponding plan is. Because of sigmoidal shape and the range, we assume that the probability to adopt one from two systems depends on the sine of difference of $e_S(X)$ and $e_L(X)$ as follows :

$$p_L(X) = 0.5\{\sin(e_L(X) - e_S(X)) + 1\}, \quad (2)$$

$$p_S(X) = 0.5\{1 - \sin(e_L(X) - e_S(X))\}. \quad (3)$$

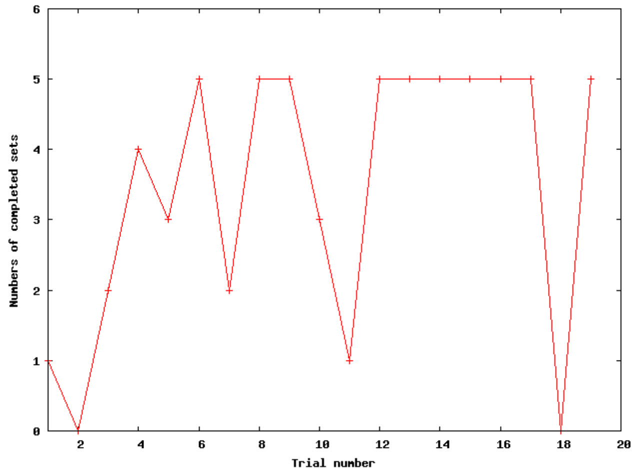
The difference of $e_L(X)$ and $e_S(X)$ is limited to the range [$-\pi/2, \pi/2$]. These values are updated within this range.

3. COMPUTER SIMULATIONS

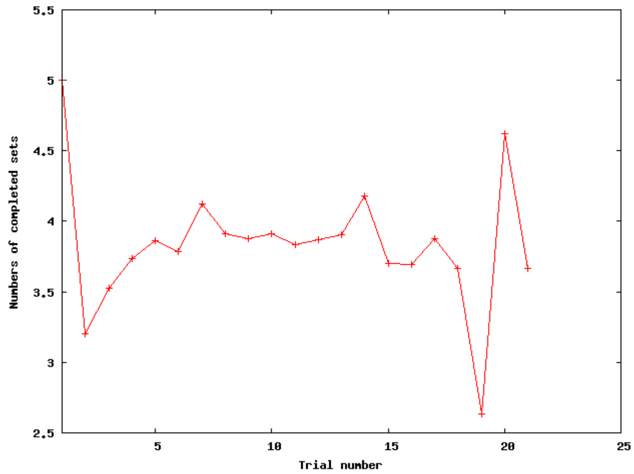
We show the results of computer simulations of the model in the new task, the learned task, and the modified task, respectively. The parameter values are selected to field results correspondent to physiological data as follow: $p_{set}(0) = 0.5$, $\gamma = 0.4$, $\alpha_{MAX} = 0.91$, $P_l = 0.99$, $u_1 = 0.01$, $u_2 = 0.1$, $u_3 = 0.5$, $u_4 = 0.1$, $u_5 = 0.01$, $b = \frac{\pi}{2}$.

3.1 New task

During this task, no stimulus set is assumed to be memorized in LTMS. As the learning proceeds, it gradually becomes to act correctly because of the role of STMS. **Fig.2** shows a typical example of temporal change of the number of successful sets. The average number of mistakes until 10 trials successful is 10.2±6.0 (which is close to 10.3±5.9 from experimental data [1]). In this task, STMS was almost selected in DSS because novelty of the first stimulus decreases e_L . This activated STMS may correspond to higher activities in Pre-SMA in this task of experimental data [1][4][5].



(a)



(b)

Fig.2. Simulation results in the new task: (a): Numbers of completed sets. (b): Numbers of completed sets from the first successful trial averaged over 100 tasks.

3.2. Learned task

In this task, the familiar stimulus in the first set increases e_L greatly and immediately. The probability of selection of STMS has been kept small, because the probability for success by STMS is much lower than that by LTMS. Fig.3 shows the examples of numbers of successful sets in the learned task. It is shown that the model performed perfectly for already learned hyperset. The average number of mistakes until 10 trials successful is 0.48±0.72, (which is close to 0.5±1.2 from experimental data [1]). In this task, almost all plans proposed by LTMS were adopted by DSS.

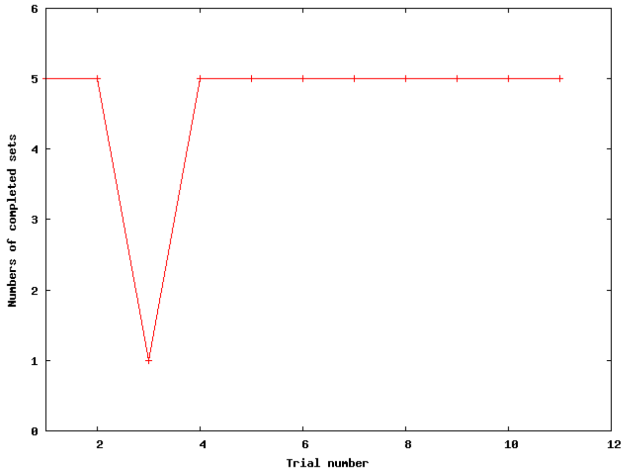
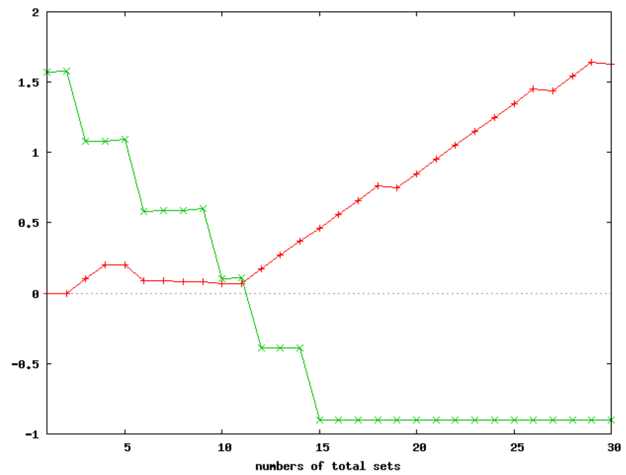


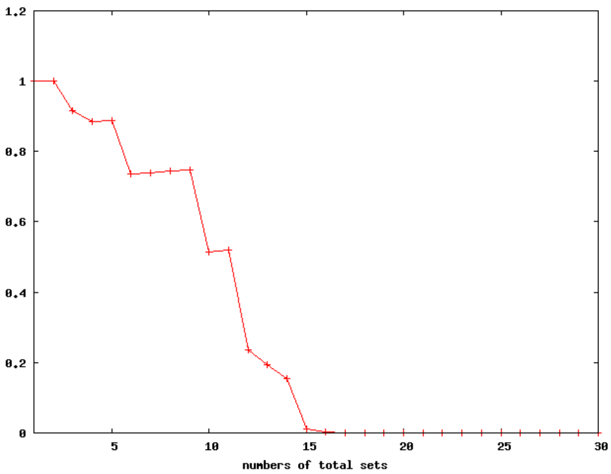
Fig.3. Simulation results in the learned task: Numbers of completed sets.

3.3. Modified task

In this task, second and third stimulus sets in hyperset are modified. Table2. shows the actions and results for each stimulus set. The meaning of abbreviations in table2. are as follows. SC: correct result by STMS plan, SW: wrong result by STMS plan, LC: correct result by LTMS, and rC/rW: correct or wrong result by retried STMS plan after ineffective LTMS action. In the early trials of the task, DSS selected the plan from LTMS because the first set is a learned one. In table2., trial #1 and set #2, trial #3 and set #2, and trial # 4 and set #2 indicate the case, when the subject pushed an unilluminated button in vain and retry another strategy of STMS. When the errors in response to the modified sets occur, the selected plan in DSS will be replaced by STMS with decreased e_L . At the same time, the correct rate of the plan from STMS gradually increases following eq.(1) from 50% as the learning proceeds. If the correct probability increases, its estimate increases e_S . Fig.4 shows temporal change of the values of e_L and e_S . Finally the performance was improved with high accuracy rate adopting the STMS plan for all of the sets.



(a)



(b)

Fig. 4. (a):The temporal change of weight parameters. (Green line denotes e_L and red line denotes e_S) (b):The temporal change of the probability of selection of LTMS

Table2. Actions and results in the modified task

Trial number	Set number 1	2	3	4	5
1	LC	rC	SC	LC	LC
2	LC	SW			
3	LC	rW			
4	LC	rC	SC	SC	SW
5	SC	SC	SC	SW	
6	SC	SC	SC	SC	SC
7	SC	SC	SW		
8	SC	SC	SW		
9	SC	SC	SW		
10	SW				
11	SC	SW			
12	SC	SC	SC	SC	SC
13	SC	SC	SC	SC	SC
14	SC	SC	SC	SW	
15	SC	SC	SC	SC	SC
16	SC	SC	SC	SC	SC
17	SC	SW			
18	SW				
19	SC	SC	SW		
20	SC	SC	SC	SC	SC
21	SC	SC	SC	SC	SC
22	SC	SC	SC	SC	SC
23	SW				
24	SC	SC	SC	SC	SC

4. CONCLUSIONS

We propose a simple action selection model in the brain, which deals with pushing two illuminated buttons in the correct order. In this model, there are two independent action-planning systems using either of long-term memory or short-term memory (working memory), respectively. When the stimulus set is input, they propose action plans, and the selection is decided in a competitive way via the value of estimation parameters. The results show that our model could make a few errors to push an unilluminated buttons following long-term memory at first but it gradually learns correct responses by trial and error using working memory in the modified task, as the monkey did in the physiological experiment. In addition, the model also reproduces similar behaviors to the biological data in the new and learned task.

On the other hand, the value of the reliability parameter e_L and e_S show common tendencies with the neuronal activities of SMA or pre-SMA in the previous experiments [1][3], in that some neurons in the SMA are relatively activated in the learned task and some pre-SMA neurons are activated in the modified task, respectively. In the modified task, the value of e_S increases not only for the novel stimulus but also for the learned stimulus as the observed pre-SMA neurons. If the distinctive memory system in our model actually works in the brain, SMA and pre-SMA may have close relations to the entries to long-term memory and working memory systems in the brain, respectively.

To confirm this hypothesis, more refinements of the model plausible to biological data and computer experiments are needed in our future works. Other issues include how to implement interactions between decision system of strategy and memory systems, and memory consolidation from the short-term memory to long-term memory.

REFERENCES

[1] Nakamura K, Sakai K, Hikosaka O (1998), Neuronal activity in medial cortex during learning of sequential procedures.*J.Neurophysiol*.80, 2671-2687

[2] Nakahara H, Doya K, Hikosaka O (2001), Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences - a computational approach. *Journal of Cognitive Neuroscience,* 626-647

[3] Sakai K, Kitaguchi K, Hikosaka O (2003), Chunking during human visuomotor sequence learning. *Exp Brain Res* 152, 229–242

[4] Nakamura K, Sakai K, Hikosaka O (1999), Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. *J. Neurophysiol*.82, 1063-1068

[5] Hikosaka O, Nakamura K, Sakai K et al(2002),Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*.12, 217-222