Abstract—We discuss about efficient motor babbling under the example of drawing learning scenario. Motor babbling is infant's motion in that infant learns the relationship between its joint input and motion. In previous work, motor babbling has been implemented to humanoid robots as random exploration processes. However, the motor babbling of human is not a completely random process in authors' insight. Motor babbling is featured by exploration and exploitation processes. The previous motor babbling processes that try only exploration is not efficient. We propose two motor babbling models, exploitation babbling and $\epsilon$-greedy babbling. In order to implement exploitation babbling, we developed an online incremental function approximator, dynamics learning tree (DLT). DLT showed 2.23 times smaller learning error and 590 times smaller computational time than neural network. Exploitation babbling successfully realized constraint in a babbling process. $\epsilon$-greedy babbling converged its prediction error fastest among the three babbling models. In the drawing that is based on Scribbling and Fortuitous Realism of Louquet, $\epsilon$-greedy babbling showed the best performance among them.

I. INTRODUCTION

In recent years, motor babbling is actively researched in the field of constructive research for the cognitive development, which focuses on the mechanism of the recognition and behavior architectures of humans [1], [2]. Motor babbling is the motions through that infants learn their body dynamics models. In that, infants try to sample the relationship between their joint input and resulting motion. Schillaci et al. modeled the dynamics of a robot arm [3]. Saegusa et al. developed a humanoid robot that learns reaching motions through motor babbling [4]. Dawood et al. and Grimes et al. developed robots that imitate the motion of others through motor babbling [5], [6]. Especially, Mochizuki et al. developed a humanoid robot that learns the dynamics of its arm and a grabbed pen using Tani's neural network model [7], [8]. The performance of this model was improved by applying stopping/pausing sequences in the training data [9]. This theory is supported by several researches [10], [11]. Watanabe et al. proposed online incremental drawing learning through body babbling [12]. In these researches, motor babbling is considered as a random exploration process that does not include exploitation concept. In this paper, we propose a model that combines exploration and exploitation processes.

In the field of reinforcement learning exploration and exploitation are frequently discussed [13], [14], [15], [16]. However, its implementation and effectiveness in motor babbling have not been researched. In our insight, an online incremental learning process is suitable for this implementation, because learning system must be updated simultaneously with a motor babbling process. Thus, we developed the online incremental function approximator named dynamics learning tree (DLT) in order to implement the proposed models.

In this paper, we propose two motor babbling models, exploitation babbling and $\epsilon$-greedy babbling. Exploitation babbling realizes the characteristic of exploitation in motor babbling. $\epsilon$-greedy babbling combines exploration babbling (previous babbling) and exploitation babbling seamlessly using the concept of $\epsilon$-greedy method. The definitions of exploration and exploitation babblings are as follows:

1) Exploration babbling (previous babbling): Random babbling that does not require any knowledge of self body.
2) Exploitation babbling: The babbling that requires the knowledge of self body. This babbling is constrained by some self body knowledge.

In this paper, we show the effectiveness of $\epsilon$-greedy babbling.

We validated the proposed models in drawing task. From the engineering approach, drawing task is tried using several methods. Kudoh et al. developed a robot system that extracts 3D model of a real object using a stereo camera [17]. The extracted model is drawn by a robot arm. Kulvicius et al. proposed a way to join motion primitives. Motion primitives are modified to highly precise trajectories [18]. Yokoyama et al. developed a motion sensing robot system that captures professionals’ drawing motion. The captured motion was regenerated by a robot arm [19]. Compared with motor babbling approach, these methods more depend on priori knowledge of the tasks and their own platforms. In order to relax the dependency, we tried this task based on the approach of motor babbling.

As Louquet [20] mentioned, drawing motion development is illustrated with following 5 phases: Scribbling (1-3 years), Fortuitous Realism (2-4 years), Failed Realism (3-7 years), Intellectual Realism (4-8 years), and Visual Realism (8+ years). In Scribbling, infants try motor babbling in order to learn the relationship between their joint input and drawn lines. They do not concern about the meaning of the lines. In Fortuitous Realism, children gradually understand the relationship between real objects and drawing. They enlarge their imitative motivation. In Failed Realism, children make a push to draw a real object. But, it fails because of insufficient capability. In Intellectual Realism, children draw imaginary things. In Visual Realism, perfect imitation is done. In this paper, we focus on Scribbling and Fortuitous Realism. Although this developmental process is researched in [7], [9], [12], it is
not completed. In this research, the characteristics of 'online incremental learning' and 'simultaneous learning for constraint and motion' are added as a new contribution for the drawing task. Also, the learning target was extended from 2 degrees of freedom (DOF) to 5 DOF.

The structure of this paper is as follows: The proposed babbling models are described in Section 2. The developed online incremental learning system, DLT, is explained in Section 3. The experimental settings and results are shown in Section 4. The discussions about the key topics are in Section 5. We conclude this paper in Section 6.

II. PROPOSED BABBLING MODELS (EXPLOITATION BABBLING AND \(\epsilon\)-GREEDY BABBLING)

In the first stage of human development, motor babbling is used for their motion learning. In previous work, motor babbling has been considered as a random exploratory process, in which a variety of joint inputs are explored randomly.

However, after an infant grows to a child, his/her behavior is not just random. For example, when a child is trying to open a bottle, the child may try random manipulation around the bottle. But, in this case, motor babbling is not completely random in joint input level, because he/she keeps his/her hand close to the bottle. For another example, when a child is trying to draw something on a canvas, the child may try to move a pen randomly on the canvas. In this case too, the children’s joint input is not completely random, because the pen tip position is controlled on the canvas. Thus, there is a transition from random exploration to ordered exploration in the developmental process of human. In the above mentioned examples, the child is exploiting his/her knowledge in order to keep his/her hand/pen tip position around the bottle/canvas, while keeping the exploration (random babbling) of manipulation/drawing motion. In previous work, this type of transition has not been discussed yet.

In order to realize the transition, we propose two babbling models, exploitation babbling and \(\epsilon\)-greedy babbling. Especially, in \(\epsilon\)-greedy babbling, motor babbling is featured by exploration and exploitation behaviors. Fig. 1 shows the flow of the previous babbling, which is featured by exploration. In this babbling, joint input is decided at first. After taking the action using the input, the relationship between the joint input and action result is learned. In this babbling, the exploration is done without any constraint. Fig. 2 shows the proposed exploitation babbling (Proposed Babbling 1). In this babbling, after joint input is decided, its result is predicted by an online incremental learning system. According to the prediction result, taking the action or not is judged. If it is accepted, the action is taken. Finally, the result of the action is feedbacked in order to improve the prediction of the online incremental learning system. In previous work [7], [9], neural network was used for the learning system. However, neural network is not suitable for online incremental learning, because of forgetting problem [21].

The exploitation babbling is able to keep constraints by rejecting the prediction results that do not meet the constraints. But simultaneously, this babbling makes sticky actions around predictable body states, because it rejects the input that does not result in a good prediction in high possibility. Therefore, this model is considered as an exploitation process.

In the theory of reinforcement learning, \(\epsilon\)-greedy method has been used for balancing the exploration and exploitation processes [13]. According to the theory of \(\epsilon\)-greedy method, the balance of exploration and exploitation processes are adjustable by selecting one of these processes using probability value \(\epsilon\). However, its implementation and effectiveness have not been researched in motor babbling.

III. DYNAMICS LEARNING TREE

In order to implement Proposed Babbling 1 and 2, we need to employ an online incremental learning system. In previous researches, neural network was employed for the learning. However, supporting online incremental learning using NN is difficult because of the problem of forgetting.

French et al. showed the forgetting problem (catastrophic forgetting) that occurs in online (incremental) learning of neural network [21]. Catastrophic forgetting is the phenomenon in that already learned knowledge on NN disappears rapidly by the online incremental learning of new training data with BP method. French et al. showed the actual example that a small weight change on a synapse erases a large amount of learned knowledge from NN. Ans et al. and Robins et al. proposed consolidation learning (CL), which applies BP method after mixing new and pseudo training data obtained from NN [22], [23]. CL is effective to avoid catastrophic forgetting. However, forgetting problem in online incremental learning is not completely avoided.

Therefore, we developed a new learning system named dynamics learning tree (DLT) [24], [25], which is able to learn new training data with the same learning weight as that of previously learned data one by one. Its online incremental learning is supported by statistics.

Dynamics learning tree is a tree typed multi-layered learning system. Fig. 3 shows the example of DLT with \(N\) layer 2 sub-layers (dimensions) 2-ary tree. DLT’s root node corresponds the whole region of \(n\) dimensional input space. Each main layer has \(n\) sub-layers (dimensions) with \(d\)-ary. The leaf nodes represent the divided sub-space as numbers 1-4 in Fig. 3.

DLT learns the relationship of continuous I/O functions. The example of its learning is shown in Fig. 4. When input data
The tree type data structure of DLT (left) represents the division of input space (right). In the Layer 1, assigned numbers 1-4 in the tree correspond the numbers 1-4 of the divisions of the input space. In this figure, all of the branches are illustrated, although these branches are incrementally created in the actual learning process.

Fig. 3. DLT (N Layer 2 Dim. 2-ary).

is given as the circle, a sub-space is created according to the input, so that the input data is in the sub-space. According to the position-on-the-tree of the node that represents the sub-space, A sequence of nodes are created from Root. Resulting tree of DLT is shown in Fig. 4 upper left. In these figures Cell 1 and Cell 2 are corresponding to Node 1 and 2 respectively.

In every node of DLT, an average output vector was retained using following update functions:

\[
\hat{O}_{\text{Cell } n} \leftarrow (N_{\text{Cell } n} \times \hat{O}_{\text{Cell } n+O})/(N_{\text{Cell } n}+1)
\]

(1)

\[
N_{\text{Cell } n} \leftarrow N_{\text{Cell } n} + 1
\]

(2)

where \(N_{\text{Cell } n}\) is the learned number of output data in Cell \(n\), \(\hat{O}_{\text{Cell } n}\) is the average vector learned by Cell \(n\), \(O\) is training output vector. DLT cancels Gaussian noise around the training output vectors by the averaging process. These equations are characterized by online incremental update processes statistically. Also, its averaging process assures the same weight of learning between previously and currently learned training data. This update function is applied to a sequence of nodes from Root to the leaf node that corresponds to training output vector \(O\). Thus, all of the nodes retain the average output vectors.

In the prediction process of DLT, DLT returns the averaged output vector of the node that represents the partition of the input space in that input data is placed. Thus, when the number of learned training input data is sparse around the new input data, DLT’s output is calculated using shallow nodes in the tree. On the other hand, when it is dense, deep nodes are used. In intuitive understanding, DLT is controlling its complementation method against a new input according to the density of the training input data.

IV. EXPERIMENT

We conducted three experiments. In Experiment 1, the learning capability of DLT was compared with NN for the validation. In Experiment 2, the three babbling models are compared in order to validate the effectiveness of the proposed babbling models. In Experiment 3, above mentioned Scribbling and Fortuitous Realism of Louquet [20] was examined using the three babbling models.

A. Settings

1) Experiment 1: We validated the learning capability of DLT with simple-harmonic-motion (SHM) learning task. For SHM, we used \(\dot{q} = -q\).

DLT with 6 layer 2 dim. 3-ary was employed. NN with three layers and BP method was employed for the comparison. For both learning systems, their input and output are set at \((q, q, \dot{q})\) and \(\ddot{q}\). For NN, the learning rate and the number of middle layer nodes are preliminary tested before the comparison. We selected the best of them according to the error after 1000000 steps of learning.

The learning of DLT was done by online incremental learning, in which training data is learned one by one. The learning of NN was done by batch learning, in which 2000 training data was learned in a learning step.

2) Experiment 2: We validated the proposed babbling methods using 5 joints of the right arm of a humanoid robot NAO (Fig. 5) in a drawing task. Using a pen tablet, we obtained the position and pressure of the pen grabbed by NAO.

We obtained pen pressure values from the pen tablet in the range of [0,1] (0: without pressure). The value is 0 when the pen is far from the tablet. On the pen tablet, 1 pixel is about 0.25 [mm].

DLT with 6 layer 5 dim. 3-ary was employed. The input and output of DLT were set at 5 joint angles and pen tip information respectively. The pen tip information is composed of \(x, y\) coordinate on the pen tablet and pen tip pressure. We made two categories for the pen tip data in order to implement the judgment function of exploitation babbling. The first is effective data that was sampled when pen tip is on the tablet. The second is ineffective data that was sampled out of the tablet. When pen tip position was \(y > 500\) or pen pressure is 0 or 1, the input was rejected. The data with 1 of pen pressure is rejected exceptionally because, pen is too strongly fitting on the tablet in this state. Also, for the initialization of the learning system of the proposed babbling models, 100 data was sampled using previous babbling model. For the \(\epsilon\) of Proposed Babbling 2, we used 0.5.

3) Experiment 3: In order to realize above mentioned Scribbling and Fortuitous Realism, which is the third phase of Louquet, we conducted an imitation experiment. A figure was copied by the humanoid robot NAO.
The original figure was given by the experimenters. We generated the data of the original figure using the technique of stopping/pausing sequences [9]. The predicted figures were generated by the rehearsal of DLT that was trained in Experiment 2. The procedure of this rehearsal is as follows:

1) DLT got a target pen tip position from the original data
2) 1000 sets of random joint angles were inputted to DLT in order to predict resulting pen tip state.
3) A set of joint angles that give the most close pen tip position as the target was selected from the 1000 sets.
4) Go to 1) again.

Figures were obtained from the actual drawing of NAO. In the drawing the joint angles that were calculated by the rehearsal were used as target joint angles of the robot. In the actual implementation, the rehearsal and drawing are done simultaneously. Thus, the prediction is included in the online control process of the robot.

B. Results

1) Experiment 1: In Figs. 6 (1) and (2), learning results of DLT are shown. After 100 data was learned (Fig. 6 (1)), the answer and prediction are not close still. After 5000 data was learned (Fig. 6 (2)), they are almost the same. Prediction errors of DLT and NN are plotted in Fig. 7. DLT converged its error much faster than NN.

In Table IV-B1, the learning results of NN after 1000000 steps are listed. The best result of NN was compared with DLT in Table II. From the result, DLT converged to 2.23 times smaller error with 500 times smaller computational time than NN.

2) Experiment 2: The numbers of sampled effective and ineffective data are listed in Table III. Proposed Babbling 1 and 2 sampled a large number of effective data compared with previous babbling.

The pen tip positions that were used for the sampling is plotted in Figs. 8. Proposed Babbling 1 generated the sampling positions with a bias. The region of data distribution in Fig. 8 (2) lacks upper left part compared with the others.

By using 1268 test data sampled by previous babbling, the prediction errors of the three babbling models are validated. For the test data, the data with pen pressure in the range of (0,1) was used. The prediction errors are plotted in Figs. 9–11. In this validation, Proposed babbling 2 converged fastest.

3) Experiment 3: Figs. 12 (1)-(3) show the original, predicted, and actually drawn figures of three learning models. Figs. 12 (4)-(6) show the errors between original & prediction, prediction & drawn, and drawn & original of them. Table IV shows the average prediction errors that were calculated from the data of Figs. 12 (4)-(6).

From Figs. 12 (4)-(6), the error between original & prediction are not so large, but the others. Even if the learning...
models' prediction is on the original figures, their actual results have large errors in some cases. Among the three models, Proposed Babbling 2 showed smallest error between drawn & original. This means that the prediction of Proposed Babbling 2 is more trustable than the others. As in Figs. 12 (1)-(3), the drawn lines of Proposed Babbling 2 is closest to the original.

V. DISCUSSION

A. Validation of DLT

From Experiment 1 and 2, DLT is applicable to the mapping of continuous input and output of dynamical systems. Also, DLT realized 2.23 times smaller error with 1/500 calculation effort than NN. The reason of this result might be because DLT adaptively tunes its generalization according to given data samples using an adaptive tree. In the learning of NN, the volume of the input space region a set of training I/O affects to the corresponding output is not governed with a theoretical background. On the other hand, DLT adaptively tunes the affection according to the density of data.

B. Validation of exploitation babbling

From Table III, exploitation babbling (Proposed Babbling 1) sampled the largest number of the effective data in that pen tip position is on the tablet. The number of ineffective data was about the half of that of exploration babbling. Proposed Babbling 1 successfully learned how to keep the constraint that was defined by the judgment process simultaneously with maintaining babbling.

C. Validation of $\epsilon$-greedy babbling

From the results of Experiment 2 and 3, the learning performance of $\epsilon$-greedy babbling always exceeded the other babbling models. This reason might be as follows. Exploration babbling does not sample data efficiently, because it does not keep constraint. Thus, it wastes the number of sampling trials to sample not required data that is out of constraint. Exploitation babbling samples data on the constraint. However, it sticks around the confidential pen tip positions that were already learned. As a result, exploitation babbling lacks its exploration ability, and makes lack of data like in Fig. 8 (2). $\epsilon$-greedy babbling relaxes the sticky tendency of exploitation babbling by adding exploration babbling probabilistically. When exploitation babbling is done, pen tip position is almost on the constraint. Small exploration around the position results in a good exploration that almost keeps the constraint. Thus, in $\epsilon$-greedy babbling, explored region does not expand too far from the constraint like exploration babbling.

D. Similarity with Scribbling and Fortuitous Realism

As in Experiment 3, Proposed Babbling 2 ($\epsilon$-greedy babbling) realized better imitation than the other models. We could confirm the following similarity with above mentioned Scribbling and Fortuitous Realism of Louquet [20]. In Scribbling and Fortuitous Realism, children learn how to move their
body through drawing trials. This learning is characterized by online incremental improvement of their prediction for pen tip movement (1. online incremental learning). Also, before the learning, children do not know even how to keep constraint of the canvas (2. inexistence of pre-knowledge for the task). The constraint keeping and motion prediction are learned simultaneously (3. simultaneous learning for constraint and motion). Once knowledge for the prediction is learned, the knowledge is used for the imitation task that is first experience without additional learning (4. acquired knowledge is applicable to a variety of tasks). Also, even in the first experience task, Proposed Babbling 2 can improve its prediction from the sampled data using online incremental learning process. Especially, the characteristics of 1 and 3 have not been realized in previous babbling models [7], [9], [12].

E. Extending DLT to neural network model

The structure of DLT might be extendable to a neural network model. The update of DLT is characterized by the path through Root to a leaf node that is selected according to the input. In the upper layer that is close to Root, a cell represents a large volume of input space. On the other hand, in the lower layer that is close to a leaf node, a cell represents a small volume of input space. When upper layer neurons gate the input that activate lower layer neurons, neural network can obtain the same type of hierarchy on its multi layered structure. Also, the update function of a cell of DLT is characterized by the decay of the update rate. This kind of decay can be modeled by the decay of the plasticity of a neuron.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed two body babbling models, exploitation babbling and $\epsilon$-greedy babbling. These babbling models are implemented using DLT. From the validation results, DLT realized 2.23 times smaller error with 1/500 calculation effort than NN. Proposed Babbling 1 successfully learned how to keep the constraint that was defined by the judgment process simultaneously with maintaining motor babbling. Learning performance of $\epsilon$-greedy babbling always exceeded the other babbling models. Using $\epsilon$-greedy babbling, the drawing learning that was closer to Scribbling and Fortuitous Realism than previous researches was realized.

The theory of $\epsilon$-greedy babbling has much potential to be applied to a variety of tasks and a variety of robots.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Numbers, 16H05877, 15K00363, 15K20850, 24119003, and SCAT Technology Research Foundation.

REFERENCES