Collaboration Development through Interactive Learning between Human and Robot

Tetsuya OGATA[1)2)] **Noritaka MASAGO[2)]** **Shigeki SUGANO[2)]** **Jun TANI[1)]**

1) RIKEN
Brain Science Institute
2-1, Hirosawa, Wako-shi, Saitama, Japan
E-mail: ogata@brain.riken.go.jp

2) Waseda University
Humanoid Robotics Institute
3-4-1, Okubo, Shinjuku-ku, Tokyo, Japan
E-mail: ogata@paradise.mech.waseda.ac.jp

Abstract

In this paper, we investigated interactive learning between human subjects and robot experimentally, and its essential characteristics are examined using the dynamical systems approach. Our research concentrated on the navigation system of a specially developed humanoid robot called Robovie and seven human subjects whose eyes were covered, making them dependent on the robot for directions. We compared the usual feedforward neural network (FFNN) without recursive connections and the recurrent neural network (RNN). Although the performances obtained with both the RNN and the FFNN improved in the early stages of learning, as the subject changed the operation by learning on its own, all performances gradually became unstable and failed. Results of a questionnaire given to the subjects confirmed that the FFNN gives better mental impressions, especially from the aspect of operability. When the robot used a consolidation-learning algorithm using the rehearsal outputs of the RNN, the performance improved even when interactive learning continued for a long time. The questionnaire results then also confirmed that the subject's mental impressions of the RNN improved significantly. The dynamical systems analysis of RNNs support these differences and also showed that the collaboration scheme was developed dynamically along with succeeding phase transitions.

1. Introduction

Recently, studies about welfare robots or pet robots, whose purposes are to actualize flexible and cooperative work with humans has attracted much attention. A humanoid robot, for example, will not only have to help people work but also have to establish a new relationship with people in daily life.

We focused on interactive learning between a human operator and a robot system, in a fundamental form to design a natural human-robot collaboration. It consists of the robot system, which learns the task including a human operator, and the human, who learns the task including the robot system. However, it is usually difficult to stabilize the system for a long period of time of operation because the incremental learning of such coupled and nested systems between humans and robots tends to generate quite complex dynamics.

Although there have already been some studies of learning systems in man-machine cooperation (Hayakawa, Y. et. al, 2000) (Sawaragi, T. et. al, 1999), most of them only focused on short period operations in which the cooperative relation between the person and the machine is organized. Therefore, they did not discuss important aspects such as the mutual interaction after the relation organization, the collapse and modification of the relation, and the long process of development from a beginner to an expert.

Miwa et. al performed an experimental study (Miwa, Y. et. al, 2001) exploring the collapse and modification of relationships between people, but such phenomena are hard to analyze because human learning and cognitive processes cannot be measured directly. Miyake et al. studied the walking cooperation between a person and a robot model (Miyake, Y. et. al, 1999), but because of their simple modeling using a nonlinear oscillator, their analysis was limited to some simple phenomena such as the synchronization and revision of the walking rhythm.

To investigate interactive learning for a long period, we developed a navigation task performed by a humanoid robot. In this task, the interaction way does not converge to a pattern but diverge to various patterns through the learning. It is thought that the epigenetic process would be important to understand these interactions. This paper describes the results of our experiments and the validity of the "consolidation learning" method implemented to ensure the robustness of neural network output.

2. Navigation Task

A navigation task is employed in which a humanoid robot, Robovie, developed in ATR (Ishiguro, H. et. al, 2001), and a human subject navigate together in a given workspace. Robovie is a small robot as a humanoid type robot, 1200 mm in height and weighing about 60 kg. It has various features enabling it to interact with human beings: two arms with four degrees of freedom, a humanlike head with audiovisual sensors, and many tactile sensors attached to its body. Photographs of Robovie and the navigation task are shown in Figure 1. The experimental environment was a 4x4-m L shaped course whose outside walls were marked red and blue for every block (Figure 2). The robot and the human subject held their arms together and attempted to travel clockwise in the workspace as quickly as possible without hitting obstacles. The actual movement of the robot and the subject is determined by adding two motor forces; one is the motor vector determined from a neural network in the robot and the other is the subject's directional control force exerted to the robot's arms. The neural network in the robot is adapted incrementally after each trial of travel based on the travel performance. The performance is measured by the travel time period at each trial.

An interesting point of this collaboration task is that the sensory information is quite limited for both the robot and the subject. The robot can access only local sensory information such as ultrasonic sensors and a poor vision system (it only detects vague color information of its surroundings), but not for exact global position information. The subject's eyes are covered during the navigation task, however the subject is allowed to look around the workspace before the experiments begin. The subject has to guess his/her situation or position by means of the interactive force felt between the robot and his/her arms utilizing his/her memorized image of the workspace geometry. Both sides attempt to acquire the forward model of anticipating the future sensory image as well as the inverse model of generating the next motor commands utilizing the poor sensory information of different modalities from past experiences.

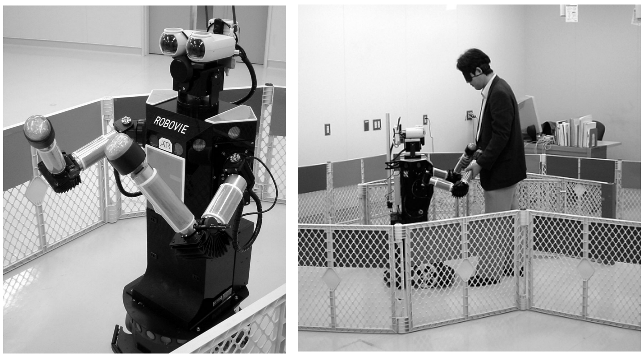

Figure 1 Robovie and Navigation Task

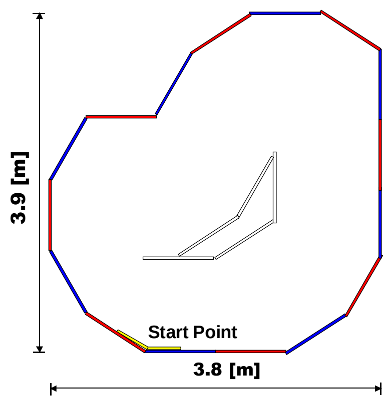

Figure 2 Experiment and Workspace

3. The Model and System

3.1 Neural Network Architecture

In many cases the actual states of the systems cannot be identified explicitly just by looking at the current sensory cues, however they do through more context-dependent manners by utilizing the history of the sensory-motor experiences. In our experiment case, the current sensory inputs may not tell the exact position of the robot due to the sensory aliasing problems. This is called the hidden state problems. Long-Ji Lin (Lin, L., and Mitchell, T., 1992) as well as Tani (Tani, J., 1996) have shown that the recurrent neural network (RNN) can solve this problem where the so-called context units are self-organized to store the contextual information. We applied the Jordan type recurrent neural network (RNN) (Jordan, M. ,1986) in which context units are added to the usual-feed forward neural network (FFNN) .

Figure 3 shows the RNN architecture design of the robot. The RNN operated in a discrete time manner with the synchronizing of each event, and the input layer of the RNN

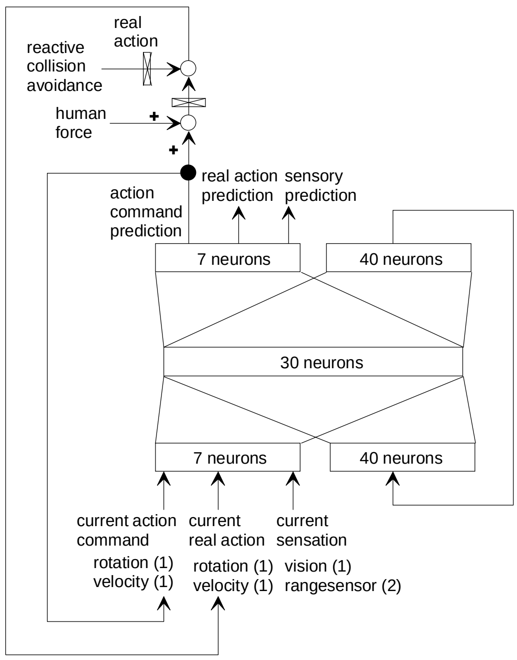

Figure 3 Neural Network Architecture

consisted of the current sensory input and the current motor values. The sensory inputs are comprised of the output of the ultrasonic range-sensors and the color area acquired from the omni-direction camera mounted on the robot's back. The motors consist of the current forward velocity and rotation velocity. The input layer has only seven units. The output layer also has seven units, and its outputs are the prediction of the next sensory input and the next motor values. This is the implementation of the paired forward and inverse model proposed by Kawato (Wolpert, D., and Kawato M., 1998). There are forty context units in the input and output layers. The activations of the context outputs in the current time step is copied to those of the context inputs in the next time step. It is noted that the context units activities are self-organized through learning processes such that they can represent the current state of the system corresponding to the past input sequences.

In our application, the RNN is utilized not only as a mapping function from inputs to outputs but also as an autonomous dynamical system. Concretely, the RNN can have two modes of operations as shown in Figure 4. The first mode is the open-loop mode where one-step prediction of the sensory-motor prediction is made using the inputs of the current sensory-motor values. The second mode is the close-loop mode in which the output re-entered the input layer through the feed back connection. By iterating this with the closed loop, the RNN can generate an arbitrary length of the look-ahead prediction for future sequences with given initial states in the input layer. This function for the look-ahead prediction of the sensory-motor sequences can achieve the mental rehearsal (Tani, J., 1998) that will be described later in the explanations of the consolidation learning. The middle layer had thirty units. The RNN was trained by using the back propagation through the time (BPTT) learning method (Rumelhart, D. et. al, 1986).

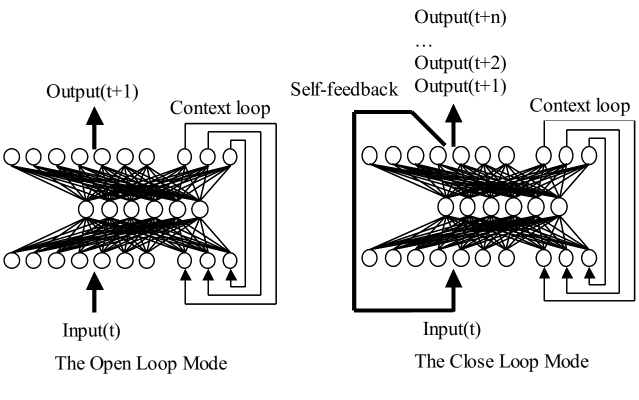

Figure 4 Open loop mode and Close loop mode

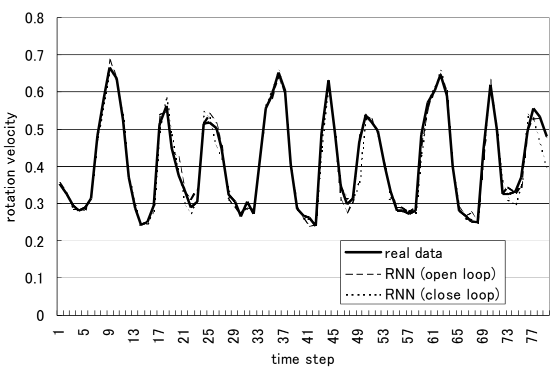

Figure 5 An Example of the result of RNN learning

Here, a pre-experimental result demonstrates how these operations of the open-loop and the close-loop modes work. Figure 5 shows the comparisons between the actual rotation velocity and its prediction in the open-loop and the close-loop modes while the robot travels in the workspace three times after the off-line training of the RNN is completed with 30,000 learning steps. From this figure, it was confirmed that the RNN had enough context neurons to memorize and reproduce the sensor data which contained noise in the real world.

3.2 Consolidation Learning

It is generally observed that if the RNN attempts to learn a new sequence, the contents of the current memory are severely damaged. One way to avoid this problem is to save all the past teaching data in a database, add new data, and use all the data to retrain the network. The problem with this method, however, is that the learning time of the RNN is increased by increasing the amount of stored data.

Therefore, we used the consolidation-learning algorithm proposed by Tani (Tani, J., 1998). Observations in biology show that some animals use the hippocampus for temporary storage of episodic memory and consolidate them into neocortical systems as long-term memory during sleep. Tani modeled this process by using a RNN and a database. In this method the newly obtained sequence pattern is stored in the "hippocampal" database. The RNN, which corresponds to the neocortex, rehearses the memory patterns, and these patterns are also saved in the database. The rehearsal can be performed in the close-loop mode described in the previous section. Various sequence patterns can be generated by setting the initial state of the RNN differently. The ensembles of such various rehearsed sequences actually represent the structure of the past memory in the dynamical systems sense. The RNN is trained using both the rehearsed sequential patterns which correspond to the former memory and the current sequence of the new experience.

It is expected that this method enables the RNN to carry out incremental learning while maintaining the structure as much as possible. Although some robot studies using this algorithm have been performed, its detailed characteristics have not yet been clarified.

3.3 Navigation system

Figure 6 shows the navigation system developed in this

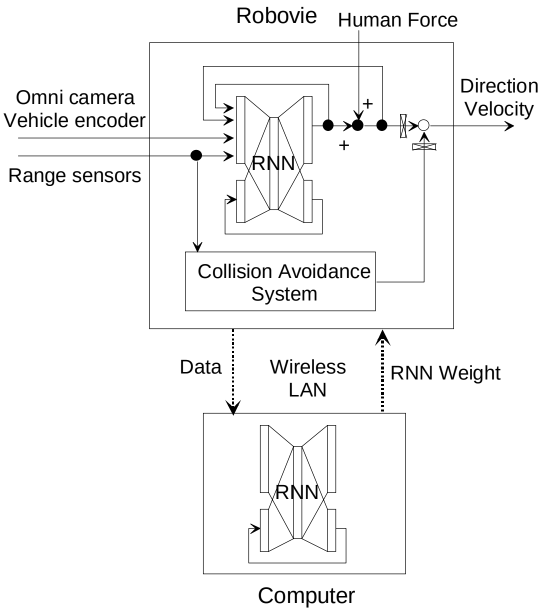

Figure 6 Navigation system diagram

study. Since the robot moves using the RNN, its performance is inadequate in the initial stage of learning. We therefore implemented a collision avoidance system which overrides the RNN commands when the minimum output of a range sensor falls below the threshold value. This system is just the reflection system tuned up by the designer. In our experiments, the more overrides made by this man-made collision avoidance system means the less performance of the RNN. The robot obtained the color area, range sensor data, and vehicle conditions every 0.1 seconds. This data was compressed and transmitted to an external computer by wireless LAN. The RNN receives this preprocessed data as input and generates the output with a time interval of 2 seconds.

A simplified reinforcement-learning method was employed for the RNN learning as follows. At each trial, the robot and the subject go around the workspace together for a fixed number of times. Then the time period taken for this travel is measured. If the performance in terms of the time period is better (less period) than the previous trial, the RNN is trained with the sensory-motor sequence experienced with this trial (with rehearsed ones in the consolidation learning). Otherwise, no training is conducted on the RNN. At each trial, 3,000 steps of iterative learning is conducted off-line using the external computer. In this way, the learning in the robot side is conducted incrementally depending on the performance achieved at each trial. Every 0.2 s the robot outputs the action commands calculated by the linear complementation of the RNN output generated every 2 seconds.

3.4 Pre-experiments

To compare the adaptability of the RNN and the FFNN, we

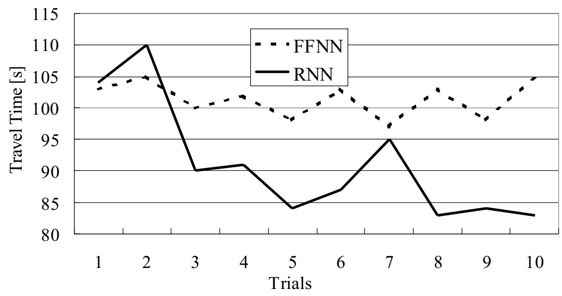

Figure 7 Performance comparison between FFNN and RNN

carried out experiments using only the robot. The RNN used in this experiment was the same as that shown in Figure 3. In this experiment we trained the RNN using the usual method rather than the consolidation method described in the previous section. The FFNN had no context layer and had a middle layer of 110 units. The total number of the neurons of both neural networks was the same (124 neurons).

Figure 7 shows the results of the 15 trials, in each of which the robot went around the workspace three times. The vertical axis shows the travel time the robot needed for one trial. It was confirmed that the RNN performed significantly better than the FFNN. This result shows that in this task the acquisition of the context information of the environment was effective. It also shows that as long as only the robot learns the environment, instability of the learning process does not cause the learning method to become a problem.

4. Experiments

The learning algorithms were evaluated and compared in 15-trial navigation experiments with seven male subjects. In each trial, the subject and the robot went around the workspace two times. After each trial there was a one-minute break for the questionnaire, which consisted of 11 items (Hayakawa, Y. et. al, 2000). The indexes were achieved by the factor analysis of 53 questions concerning 66 subjects engaged in work. Additionally, at the end of each experiment, the subjects filled out the questionnaire based on NASA-TLX (Hart, S.G. et. al , 1988). NASA-TLX is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales. It can be used to assess workload in various human-machine environments.

Three neural networks, the FFNN, the RNN with a usual learning method, and the RNN with consolidation learning explained in section 3.2 were compared in the experiments involving the seven subjects. In consolidation learning, the teaching data consisted of the current sequence pattern and the three rehearsal patterns. These three rehearsal patterns were generated in the close-loop mode with changes in the initial value of the context units randomly. Here, the number of rehearsal patterns is very crucial for the consolidation learning. In this experiment, three rehearsal patterns were enough for the consolidation, because the RNN had to memorize only one attractor corresponding to L-shaped course. The order of the experiments was changed with the subjects to avoid presenting subjects with a fixed order that might influence the results. Subjects did not know which network they were collaborating with until all experiments were finished.

4.1 Comparison of the Performances

Figure 8 shows the transitions of the travel time of three neural networks. Each travel time is the average of the seven subjects. The goal of this task was to decrease the travel time. Although all performances improved in the first half of the learning, differences appeared in the second half. The performance of the FFNN tended to deteriorate gradually and that of the RNN with the usual learning method tended to stagnate. Only the performance of the RNN with the consolidation-learning algorithm continued to improve.

The override rates of the collision avoidance system, which is equal to the operating rates of the neural networks are shown in Figure 9. These are also the average of the seven subjects. The rates showed almost the same tendency as the performances, except that the operating rate of the RNN with the usual learning method decreased in the second half of the learning. It can be considered that the subject's learning maintained the same overall performance in spite of the decrease of the operation rate of the RNN.

Through the interactive learning, the subjects changed their strategy at times. Since there were so many different ways to

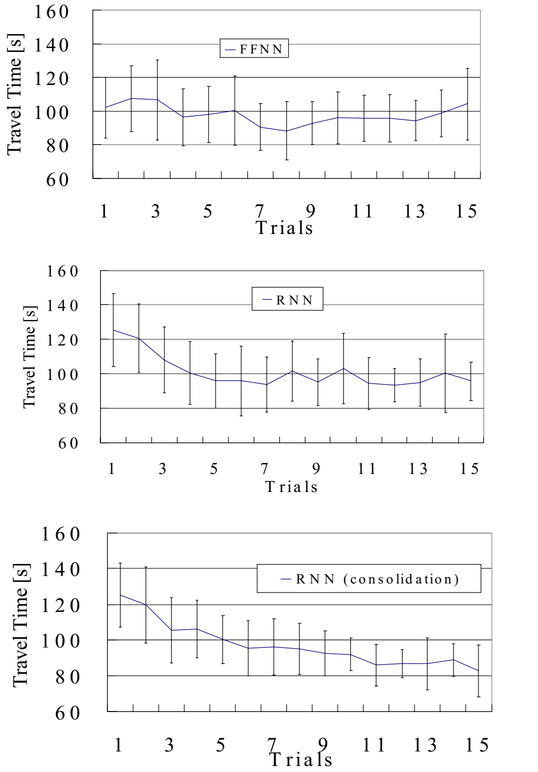

Figure 8 Performance comparison between neural networks

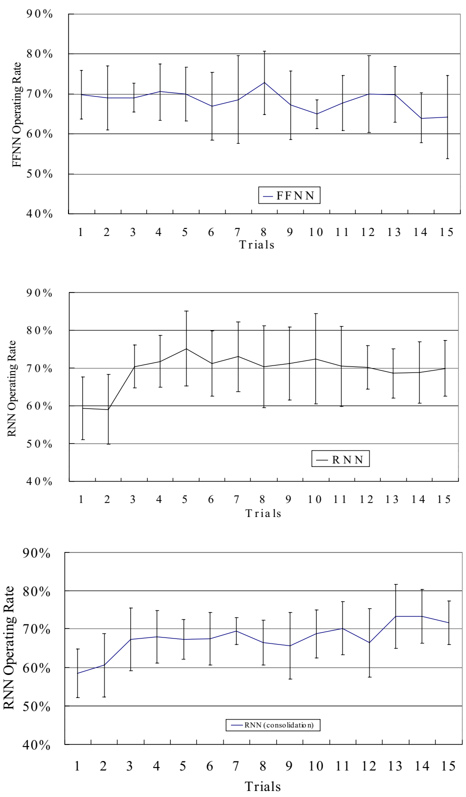

Figure 9 Comparison of operating rate of neural networks

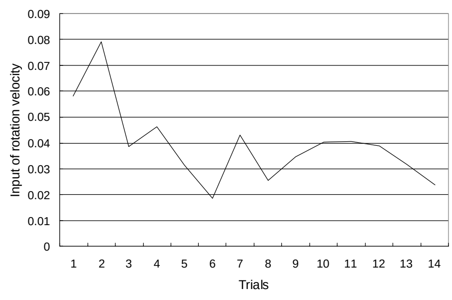

Figure 10 An Example of the Subject's Input of Rotation Velocity

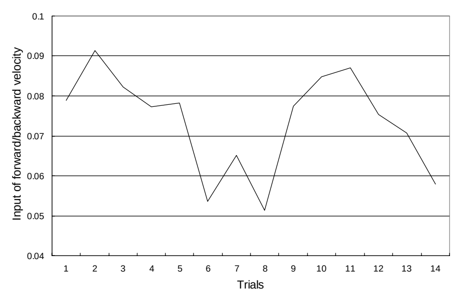

Figure 11 An Example of the Subject's Input of Forward Velocity

change their strategy, it was difficult to analyze them in general. Figure 10 and 11 show a comprehensible example of the subject's input to the robot. It was confirmed that though the input of the rotation velocity did not change so much, the input of the forward/backward velocity changed rapidly after the 7th or 8th trials. As a result, it can be said that the subject's strategy changed from the "collision avoidance" to the "speed up" in this experiment.

4.2 Mental Impression

The results of the NASA-TLX questionnaire and the 11-item questionnaire are shown in Figures 12 and 13. In each questionnaire, the significant values of the levels of 1 % and 5 % were calculated by a Scheffe test. It is easy to see that in both questionnaires the RNN with the consolidation-learning algorithm gave the best mental impressions.

It is interesting here to compare the results between the FFNN and the RNN with the usual learning method. In the robot experiments described in section 3, the performance of the RNN was better than that of the FFNN. In the navigation experiments, however, the FFNN tends to give the subjects a better impression than the RNN especially in "result of work", "fatigue free", and "operability in the 11-item questionnaire.

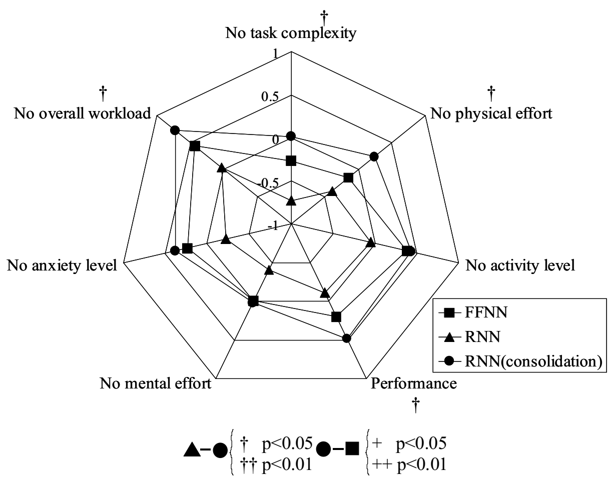

Figure 12 Results of Quesionnaire of NASA-TLX

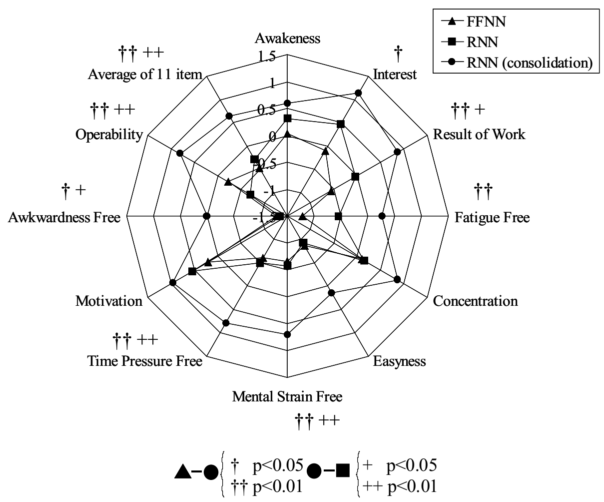

Figure 13 Results of Quesionnaire of 11 item evaluation

5. Discussion

5.1 Robustness and Operability

In the experiment with only the robot, the RNN performed effectively, because the robot could decide the action using not only the sensor input including the noises but also the information of the context layer. In the human-robot cooperation, however, the performance of the RNN with the usual learning was worse than that of the FFNN. It is thought that this is due to interactive learning including "incremental learning" which damaged the memory of the RNN. Therefore, the context layer included wrong information and created an undesired output. As a result, the "operability" became worse than that of the FFNN which has no context layer, because the robustness of the RNN to the input noise decreased.

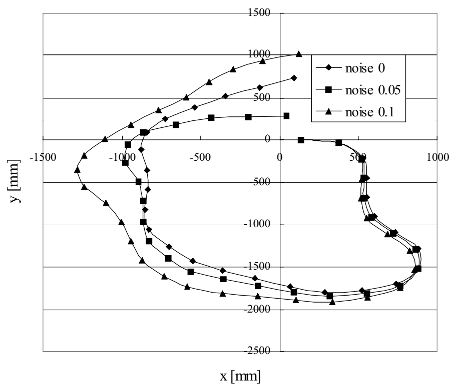

Figure 14 Examples of the trajectory with input noises
(The RNN with usual-learning method)

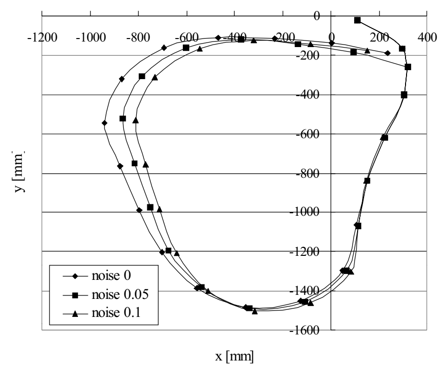

Figure 15 Examples of the trajectory with input noises
(The RNN with consolidation-learning method)

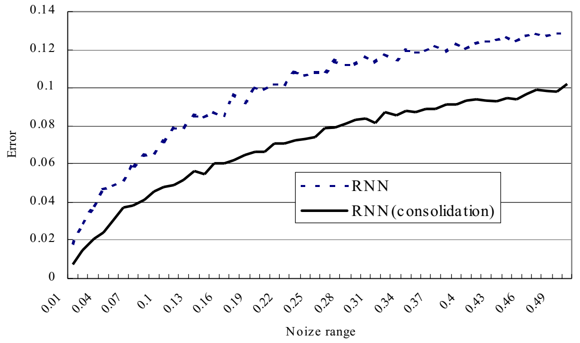

Figure 16 Robustness Comparison of RNNs

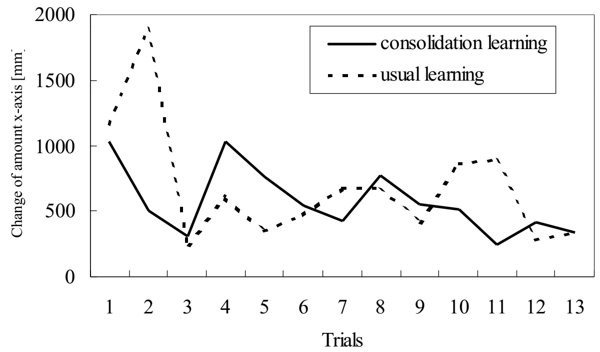

Figure 17 Transition of the Change of Rehearsal Trajectory

To analyze the effect of consolidation learning, we examined the robustness of the RNN dynamics by looking at its initial sensitivity characteristics. Both RNNs obtained after the usual learning and consolidation learning in our experiments were tested to generate the output sequences in the close-loop mode with the addition of three different sizes of noise in the initial input values. Figures 14 and 15 shows the motion trajectories of the robot re-constructed from the rehearsal motor output of the RNN. These represent how the output trajectories developed with small differences in the initial input conditions for the RNNs with usual learning and consolidation learning. It is observed that output trajectories of the RNN by the usual learning tends to diverge more than those by consolidation learning. This implies that the usual learning scheme tends to generate more unstable dynamic structures in the trained RNN.

The following analysis was carried out to investigate this property in detail. A sequence corresponding to one trial (about 100 steps) was rehearsed in the close-loop mode by the RNN. In this process, the random noises were added to seven units in the input layer for all steps. Ten sequences were rehearsed by this method. Figure 16 shows the step error obtained from the difference between the average of these ten sequences and the other sequence which was also rehearsed by the same RNN without noise. The horizontal axis is the maximum width of the noise at each step, and the vertical axis is the average error of seven RNNs corresponding to the subjects. This result also shows that the RNN with a consolidation-learning algorithm is more robust than that with the usual learning algorithm.

This robustness characteristic of the RNN seems to be directly related to the "operability" in the mental impressions. If the RNN tends to diverge largely even with small deviations in the input sequences from the learned ones in the past, it would be difficult for the subjects to harness the robot in the right directions. In the usual RNN learning, if the contents of the current memory conflict with the one to be newly learned, the internal structure of the RNN could be deflected severely and undesired pseudo memories could be generated. Consolidation learning is beneficial in this aspect since this scheme allows iterative rehearsing of past experiences. Also, training with enough number of rehearsed sequences could achieve a sort of generalization while attaining the global structures in the internal representation.

5.2 Development of Collaboration

Miwa showed that the performance of human collaboration was developed in the stages where the coherent phase and incoherent phase were repeated (Miwa et. al, 2001). He explained that this was the process in which the human generated and/or modified the hypothetical strategy for the task by using the context information.

We investigated whether the development process of the RNN would include such stages. The trajectories were rehearsed by the RNN which is the same one trained in the experiments described in section 4, and the amount of the change of each trajectory was calculated. Figure 17 shows examples of the amount of transition change of the trajectory. It was confirmed that the transition with the consolidation-learning method had three peaks (1st, 4th, and 8th trial) and decreased gradually. This means that the phase transitions occurred three times in the development process and it became stable gradually. On the other hand, the transition with the usual-learning method had no clear peak and increased gradually. This means that there was no clear phase transition in the process and it became unstable.

The trials in which the phase transition occurred in the development process with the consolidation-learning method corresponded to the trials in which the performance improved drastically. The RNN with the consolidation-learning method might have similar characteristics to humans.

6. Conclusion

In this paper, we showed that interactive learning is essential and important for man-machine cooperation. We also pointed out that it is difficult to actualize context dependence learning. The RNN was introduced as a learning algorithm system which can treat the hidden state problem. The target task was the human navigation by a humanoid robot called Robovie. The FFNN and the RNN were compared as the learning algorithm of Robovie. Although the performances of the RNN and the FFNN both showed improvements in the early stages of learning, they gradually became unstable as the subject changed the operation by learning on its own. Finally, all performances failed. The results obtained when the consolidation-learning algorithm, which uses the rehearsal outputs of RNN, was applied confirmed that this algorithm's performance was improved even when interactive learning continues for a long time and that the subject's mental impressions were better. Analyzing and comparing the characteristics of RNNs produced results which support these differences and showed that the collaboration scheme developed dynamically along with succeeding phase transitions. This process could be regarded as not just a learning process of the robot but a emergence process of various forms of mutual interaction. Consolidation learning which enable the robot to adapt the human was essential to realize this process.

Two further studies should be carried out. One should be a more detailed analysis of the characteristics of the consolidation-learning algorithm with more subjects. Although the robustness of the RNN have been compared in this paper, the relation between the dynamic structure of the RNN and the learning algorithm have not be analyzed mathematically yet. The other study that should be carried out is a transition structure analysis of the cooperation form between a human and a machine in the interactive learning process. Although we showed that the RNN with the consolidation-learning algorithm was developed with the phase transition, the adaptation phenomenon that might be the changing process of the cooperation form has not been analyzed in detail yet. Although an example which shows this changing process in section 5.2, the correspondence between the phase transition of the RNN structure and the development of human operation should be investigated through more experiments.

References

Hart, S.G. et. al (1988). "Development of NASA-TLX: Results of empirical and theoretical research." InP.A.Hancock and N.Meshkati(eds.), Human Mental Workload, North-Holland, pp. 139-183.

Hayakawa, Y., Kitagishi, I., Kira, Y., Satake, K., Ogata, T., and Sugano, S. (2000). "An Assembling Support System based on a Human Model -Provision of Physical Support According to Implicit Desire for Support," Journal of Robotics and Mechatronics, Vol.12, No.2, pp.118-125.

Ishiguro, H., Ono, T., Imai, M., Maeda, T., Kanda, T., Nakatsu, R. (2001). "Robovie: an interactive humanoid robot." International Journal of Industrial Robotics, Vol. 28, No. 6, pp.498-503.

Jordan, M. (1986). "Attractor dynamics and parallelism in a connectionist sequential machine." in Proc. of the Eight Annual Conference of the Cognitive Science Society (Erlbaum, Hillsdale, N.J), pp. 513-546.

Lin, L., and Mitchell, T. (1992). "Efficient Learning and Planning within the Dyna Framework", Proc. of the Second International Conference on Simulation of Adaptive Behavior (SAB '92), pp.281-290.

Miwa, Y., Wesugi, S., Ishibiki, C., and Itai, S. (2001) "Embodied interface for emergence and co-share of 'Ba', Usability Evaluation and Interface Design." in Proc. of HCI International 2001, pp. 248-252.

Miyake, Y., and Minagawa, T. (1999). "Internal observation and co-generative interface," in Proc. of IEEE International Conference on Systems, Man, and Cybernetics (SMC'99),I229/I-237.

Rumelhart, D., Hinton, G., and Williams, R. (1986). "Learning internal representation by error propagation", In D.E. Rumelhart and J.L.Mclelland, editors, Parallel Distributed Processing (Cambridge, MA: MIT Press)

Sawaragi, T., Kudoh, T., and Ozawa, S. (1999) "Extracting Motion Skills from Expert's Proficient Operation Records Using Recurrent Neural Network," Reprints of 14th World Congress of IFAC, Beijing, Vol.M, pp.359-364.

Tani, J. (1996). "Model-based Learning for Mobile Robot Navigation from the Dynamical Systems Perspective", IEEE Trans. on System, Man and Cybernetics Part B (Special Issue on Robot Learning), Vol.26, No.3, pp.421-436.

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

Wolpert, D., and Kawato M. (1998), "Multiple paired forward and inverse models for motor control". Neural Networks Vol. 11, pp. 1317-1329.