The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Emergence of Evolutionary Interaction with Voice and Motion between Two Robots using RNN

Wataru Hinoshita, Tetsuya Ogata, Hideki Kozima, Hisashi Kanda, Toru Takahashi, and Hiroshi G. Okuno

Abstract— **We propose a model of evolutionary interaction between two robots where signs used for communication emerge through mutual adaptation. Signs used in human interaction, e.g., language, gestures and eye contact change and evolve in form and meaning through repeated use. To create flexible human-like interaction systems, it is necessary to deal with signs as a dynamic property and to construct a framework in which signs emerge from mutual adaptation by agents. Our target is multi-modal interaction using voice and motion between two robots where a voice/motion pattern is used as a sign referring to a motion/voice pattern. To enable evolutionary signs (voice and motion patterns) to be recognized and generated, we utilized a dynamics model: Multiple Timescale Recurrent Neural Network (MTRNN). To enable the robots to interpret signs, we utilized hierarchical neural networks, which transform dynamics model parameters of voice/motion into those of motion/voice. In our experiment, two robots modified their own interpretation of signs constantly through mutual adaptation in interaction where they responded to the other's voice with motion one after the other. As a result of the experiment, we found that the interaction kept evolving through the robots' repeated and alternate miscommunications and re-adaptations, and this induced the emergence of diverse new signs that depended on the robots' body dynamics through the generalization capability of MTRNN.**

I. INTRODUCTION

Signs used in human interaction e.g., language, gestures and eye contact change and evolve in form and meaning through repeated use. Such diversity and fluctuations in signs cause miscommunication, and humans adapt to this by guessing one another's intentions. The repetition of miscommunication and re-adaptation leads to further evolution in signs. Thus, interaction is essentially evolutionary. Miwa et al. investigated the evolutionary nature of human interaction including the repetition of miscommunication and successful communication, which they called "incoherent states" and "coherent states" respectively [1].

Existing interaction robots and systems are constructed on the assumption of top-down fixed signs. While they are practical for domain-limited and goal-oriented interaction, they are not designed for flexible human-like interaction in which signs evolve.

On the other hand, agent-based methods of modeling interaction based on the viewpoint where interaction is regarded as a dynamic complex system have attracted a great deal of attention [2]. There have been several studies that have made use of these methods (e.g., Hashimoto [3], Igari and Ikegami [4]). Most of these have constructed models that dealt with highly abstracted interaction separated from the real world because their purpose was to mainly analyze the mechanisms for interaction. Therefore, it is difficult to apply these models directly to interaction robots or systems in the real world.

As a step toward achieving flexible human-like interaction robots, we construct a model of evolutionary interaction between them. In the model, diverse signs emerge on the basis of the agent robots' body dynamics and the signs are shared between them through mutual adaptation. Our target is multi-modal interaction using voice and motion between two robots, where a voice/motion pattern is used as a sign referring to a motion/voice pattern. The two robots have their own voice-motion mapping as a way of interpreting signs. They modify the mapping constantly throughout mutual adaptation in an interaction where they respond to the other's voice with motion one after the other.

There are two issues in our evolutionary interaction: first, the development of a recognition and generation system for evolutionary motion and voice signs; second, the development of a flexible voice-motion mapping system for interpreting signs. We dealt with these issues through the following approaches: first, by utilizing a dynamics model for dealing with voice and motion; second, by transforming dynamics model parameters of voice and motion mutually with hierarchical Neural Networks (NN). For the first approach, we used Multiple Timescale Recurrent Neural Network (MTRNN) [5] as the dynamics model. This model learns, recognizes, and generates sequential data through self-organizing its own parameter space with its capability for generalization. MTRNN is used to integrate the robots' sensory and motor information by learning them simultaneously; its parameter space is self-organized as a cognitive structure on the basis of the robot's body dynamics. In the second approach, the robots modify their own interpretation of signs by retraining their NN.

The rest of the paper is organized as follows. Section II gives an overview of our interaction model and the methods we utilized for it. The configurations and procedures for the interaction experimental system are described in Section III, and Section IV describes two different experiments. The first is an imitation experiment for evaluating the ability of a robot to recognize and generate signs, and the second is an interaction experiment. Section V discusses the results across disciplines. Our conclusions and future work are presented in Section VI.

Wataru Hinoshita, Tetsuya Ogata, Hisashi Kanda, Toru Takahashi, and Hiroshi G. Okuno are with the Graduate School of Informatics, Kyoto University, Kyoto, Japan {hinosita, ogata, hkanda, tall, okuno}@kuis.kyoto-u.ac.jp

Hideki Kozima is with the School of Project Design, Miyagi University, Miyagi, Japan xkozima@myu.ac.jp

978-1-4244-3804-4/09/$25.00 ©2009 IEEE

II. Interaction Model and Methods

This section gives an overview of the interaction model and methods.

In our target interaction, two robots try to convey a motion/voice pattern to the other robot using a voice/motion pattern as a sign. If the interpretation of the sign is shared between them, they have conveyed their intention to the other correctly, otherwise incorrectly.

Given the evolutionary nature of interaction, signs used in the interaction should emerge through agents' mutual adaptation, without presuming any form or meaning. So we constructed our interaction system as follows. To recognize and generate signs (voice and motion patterns), we utilized a dynamics model, MTRNN. Its generalization capability enables the robots to deal with diverse forms of signs, which are not given a priori. To interpret signs, we utilized NN, which transforms dynamics model parameters of voice/motion into those of motion/voice. Meanings of signs emerge and change while the robots retrain their own NN in interaction.

A. Interaction Model Overview

There is an overview of our interaction model in Fig. 1. An agent robot has a pair of NNs to interpret signs and two MTRNNs for voice and motion. The MTRNN for voice (Voice MTRNN) is used for two-way translation of sound waves and MTRNN parameters. The MTRNN for motion (Motion MTRNN) is used for two-way translation of robot physical movements and MTRNN parameters. The pair of NNs (Interpretation NN) is used for two-way translation of MTRNN parameters for voice and motion. A process, for example, that a robot interprets a voice sign as a motion pattern is as follows. (1) recognition: The observed voice is transformed into voice parameters by Voice MTRNN. (2) interpretation: The voice parameters are transformed into motion parameters by Interpretation NN. (3) generation: The motion parameters are transformed into a motion pattern by Motion MTRNN.

Through interchanging voice and motion, each agent robot modifies its interpretation of signs to adapt to that of the other by retraining its Interpretation NN.

B. Recognition and Generation of Signs

We utilized MTRNN for recognizing and generating signs (voice and motion). MTRNN can learn multiple sequential data and self-organize its parameter space by generalizing these data. The model is capable of learning more complex data compared to Recurrent Neural Network with Parametric Bias (RNNPB) [6], which also has similar features. The model works as a recognizer and a generator of actions by simultaneously learning sensory and motor information. Furthermore, it recognizes and generates new actions with its generalization capability. This capability provides a diversity of actions that are used as signs, which is the essence of evolutionary interaction.

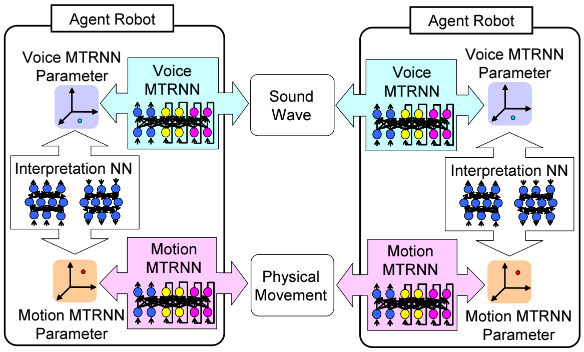

Fig. 1. Overview of Interaction Model

1) Dynamics Model MTRNN: MTRNN, proposed by Yamashita et al. [5], is an extended RNN model. Composition of MTRNN is outlined in Fig. 2. This model deals with sequential data through calculating the next state $S(t+1)$ from the current state $S(t)$. The model is composed of three neuron groups, each with an associated time constant. The three groups in increasing order of the time constant, are input/output nodes (IO), fast context nodes (Cf), and slow context nodes (Cs). The output value of the i-th neuron at step t ($y_{i,t}$) is calculated as follows.

$$y_{i,t} = \frac{1}{1+\exp(-u_{i,t})} \tag{1}$$

$$u_{i,t} = \left(1-\frac{1}{\tau_i}\right)u_{i,t-1} + \frac{1}{\tau_i}\left[\sum_j w_{ij}x_{j,t}\right] \tag{2}$$

$$x_{j,t} = y_{j,t-1} \tag{3}$$

$u_{i,t}$: internal value of the i-th neuron at step t
τ_i : time constant of the i-th neuron
w_{ij} : connection weight from the j-th neuron to the i-th neuron
$x_{j,t}$: input from the j-th neuron at step t

If neurons have larger time constant, their states change more slowly and they deal with more abstract information. Therefore, Cf represents primitives of sequential data, and Cs represents a sequence of the primitives (Fig. 3). Thus, MTRNN is able to deal with longer and more complex sequential data compared to RNNPB [6]. MTRNN represents various data patterns depending on the initial values of Cs ($\boldsymbol{Cs_0}$). Moreover the model self-organizes the parameter space ($\boldsymbol{Cs_0}$ space) through generalizing training patterns.

To train MTRNN, the Back Propagation Through Time (BPTT) algorithm is utilized. When using the algorithm, the input values $x_{j,t}$ of IO neurons is calculated with feedback from the teacher signal as follows.

$$x_{j,t} = 0.9 \times y_{j,t-1} + 0.1 \times T_{j,t-1} \quad (j \in IO) \tag{4}$$

$T_{i,t}$: the teacher signal for the j-th neuron at step t

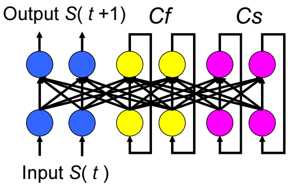

Fig. 2. Composition of MTRNN

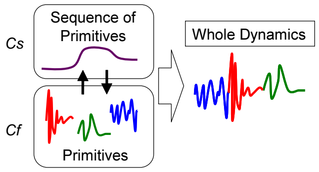

Fig. 3. Dynamics Representation of MTRNN

Connection weights (w_{ij}) and $\boldsymbol{Cs_0}$ ($x_{i,0}$ $s.t.$ $i \in Cs$) are updated as follows.

$$
\begin{aligned}
w_{ij}(n+1) &= w_{ij}(n) - \eta \frac{\partial E}{\partial w_{ij}} && (5)\\
x_{i,0}(n+1) &= x_{i,0}(n) - \alpha \frac{\partial E}{\partial x_{i,0}} \quad (i \in Cs) && (6)\\
E &= \sum_t \sum_{i \in IO} (y_{i,t} - T_{i,t})^2 && (7)
\end{aligned}
$$

n : iteration number in the updating process
η, α : learning rate constant
E : prediction error

$\boldsymbol{Cs_0}$ parameter space is self-organized depending on a dynamical structure among training patterns through the process where connection weights, which are shared by all patterns and $\boldsymbol{Cs_0}$, which is proper to every pattern are simultaneously updated. This parameter space contains various new patterns which are products of the generalization capability of MTRNN.

To recognize a sequential data, the $\boldsymbol{Cs_0}$ vector representing the data pattern is calculated through BPTT with connection weights fixed (cf. (6)). In the recognition phase, input values of IO node where the teacher signal is given are calculated by (4) and input values of the others are calculated by (3). Thus, MTRNN can recognize sequences if only partial information is given.

A sequential data is generated by executing forward calculation (cf. (1), (2), (3)) recursively with $\boldsymbol{Cs_0}$ representing the sequential data set.

2) Framework of Sensori-motor Integration with MTRNN: Each agent needs capabilities to recognize the other's signs (voice, motion) and to generate signs by itself for interaction. We enabled these capabilities with MTRNN that have learned and generalized sensori-motor sequence data obtained from each robot's actions. The trained MTRNN (sensori-motor integrated model) is capable of recognizing and generating diverse signs on the basis of each robot's body dynamics. This framework of sensori-motor integration consists of three phases: learning, recognition, and generation (Fig. 4).

1) Learning (Acquisition of Sensori-motor Integrated Model): The agent robots make actions and perceive the results of the actions through their sensors. The obtained motor-information and sensory-information sequences are learned simultaneously by MTRNN. Through this process, the MTRNN parameter space is self-organized reflecting each robot's body dynamics.
2) Recognition: An agent robot acquires the sensory-information sequence through observing the other's action. A $\boldsymbol{Cs_0}$ vector representing the action is calculated from the sensory information by BPTT with the connection weights fixed.
3) Generation: The motor-information sequence represented by a $\boldsymbol{Cs_0}$ vector is generated through the forward calculation of MTRNN.

C. Interpretation of Signs

For the interpretation of motion and voice signs, we utilized a pair of NNs (Interpretation NN) which translates $\boldsymbol{Cs_0}$ vectors for voice and those for motion mutually. To share the interpretation of signs, robots need to modify their Interpretation NN. They guess the other's interpretation through observing its behavior and modify their Interpretation NN to adapt their own interpretation to the other's. However, if their interpretations change inconsistently every time they interact, sharing signs through mutual adaptation is impossible. To maintain the consistency of each robot's interpretation after it had been modified, we utilized a consolidation learning algorithm for retraining the Interpretation NN as follows.

1) A robot obtains a voice-motion pair connected mutually by the other's interpretation through speaking to it and watching its response motion.
2) Voice-motion pairs from current interpretation are stored. To be more precise, grid-point data are input into the NN and stored with the corresponding output.
3) Interpretation NN is retrained using the new voice-motion pair obtained in (1) and the pairs stored in (2).

III. Interaction Experimental System

A. System Configuration

We used a small life-like robot "Keepon" [7] as the platform for our experiments. A Keepon has four degrees of freedom, two of which used in the experiments: PAN and TILT (Fig. 5). It also has two CCD cameras at its eyes and a microphone at its nose. The two Keepons faced each other at intervals of 230 mm, and we placed speakers beside them. The system is shown in Fig. 6.

To synthesize sound for the Keepons' voice, we used "Maeda model": the vocal tract model proposed by Maeda[8]. This model has seven parameters (listed in Table I) that determine the vocal tract shape, and sound is synthesized depending on these. By using Maeda model, we can apply the framework of sensori-motor integration to recognize and generate voices. Our method of dealing with voices is based on a study by Kanda et al. [9].

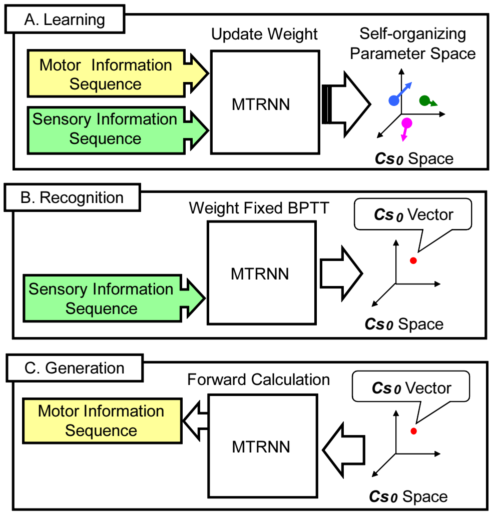

Fig. 4. Framework of Sensori-motor Integration with MTRNN

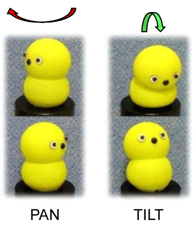

Fig. 5. Axes of Motion

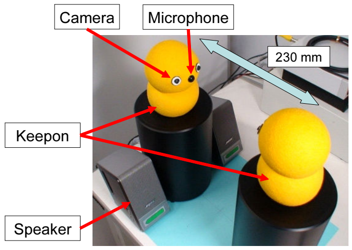

Fig. 6. Hardware Configuration

TABLE I

Maeda Model Parameters

Parameter numbers	Parameter names
1	Jaw position (JP)
2	Tongue dorsal position (TDP)
3	Tongue dorsal shape (TDS)
4	Tongue tip position (TTP)
5	Lip opening (LO)
6	Lip protrusion (LPR)
7	Larynx position (LP)

B. Procedure

The interaction emergence procedure consists of two phases. First, both robots acquire sensori-motor integrated models for voice and motion by training MTRNN. Through this process, they obtain the capabilities of recognizing and generating signs. Second, they constantly modify their interpretations of signs to adapt to those of the other by retraining their Interpretation NN throughout the interaction.

1) Phase 1 - Acquisition of Sensori-motor Integrated Model: To train Voice MTRNN, we utilized Maeda model parameters as motor information, and features extracted from the sound generated by the Maeda model from the parameters as sensory information (Fig. 7). We dealt with voice from the articulatory movements controlled by the first six Maeda parameters in Table I. We used an eight-dimensional vector of MFCC (Mel-Frequency Cepstrum Coefficient) comprising the third to the tenth coefficient as the sound features. The MFCC is configured as follows: a sampling frequency of 16000 Hz, a frame length of 25 millisecond, a frame shift of 10 millisecond, and 24 filter bank channels.

We prepared 124 articulatory movement patterns as sequential Maeda parameters composed of transitions of the four vowels /a/, /i/, /u/, and /e/. We use some of the patterns for training and rest of them to validate the ability of the trained MTRNN to recognize and generate patterns. There is an example of the patterns and sound features obtained through the voice in Fig. 8.

To train Motion MTRNN, we utilized the Keepons' motor values (PAN, TILT) as the motor information, and features extracted from the images captured by their cameras as the sensory information (Fig. 9). For the image features, a position (x, y) of an anchoring point in an image of a Keepon's eyesight was used to represent change of view according to a its physical movement. We used a center of the other Keepon's nose as the anchoring point. Sequential data of the image features obtained through watching the other's motion is the reverse of one obtained through making same motion on their own. Therefore, sensory information needs to be reversed before it is input into Motion MTRNN when a Keepon recognizes other's motion. This is a simple mechanism for transforming viewpoints. We presumed that the mechanism is given to Keepons a priori.

We prepared 80 physical movement patterns composed of combinations of PAN and TILT value sequences expressed as simple sine curves. We use some of them for training and the rest for validation. There is an example of the patterns and image features obtained through the motion in Fig. 10.

The numbers of input data dimensions for Voice MTRNN and Motion MTRNN are shown in Table II.

2) Phase 2 - Sharing of Sign Interpretation: The process of sharing sign interpretations by the two Keepons consists of the following steps. This is portrayed in Fig. 11. We call the two Keepons "**A**", and "**B**" from now on.

1) **A** speaks to **B**.
2) **B** recognizes the voice, and based on its own interpretation makes the motion associated with it.
3) **A** recognizes **B**'s motion, and retrains its Interpretation NN to connect **A**'s first voice and **B**'s response motion.
4) Repeat steps (1) to (3) but with reversed roles. (In the next sequence of the steps, the topic of the interaction is inherited.)

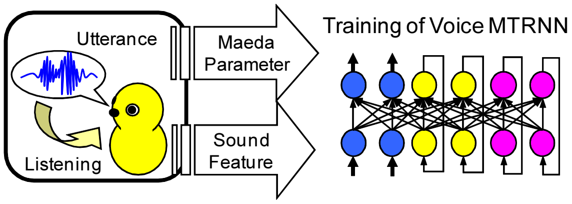

Fig. 7. Training of Voice MTRNN

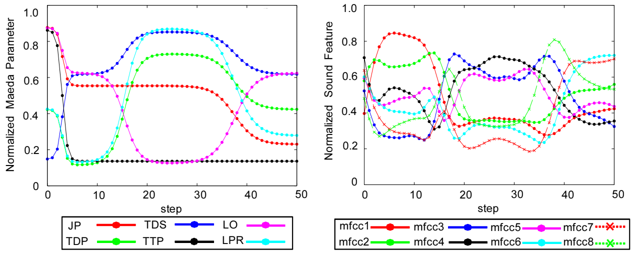

Fig. 8. An Example of Training Patterns for Voice

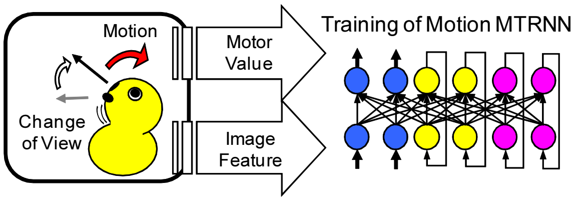

Fig. 9. Training of Motion MTRNN

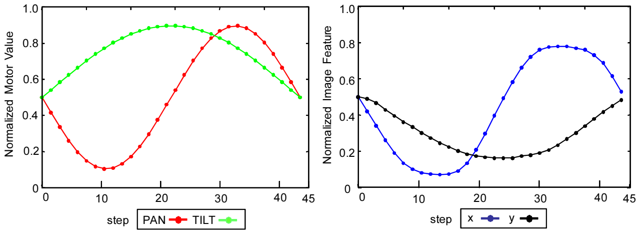

Fig. 10. An Example of Training Patterns for Motion

TABLE II

NUMBER OF INPUT DATA DIMENSIONS FOR MTRNN

	Number of Input Data Dimensions	
	Voice MTRNN	Motion MTRNN
Motor Infomation	6 (Maeda Model Parameters)	2 (Robot Motor Values)
Sensory Infomation	8 (Sound Features)	2 (Image Features)
Total	14	4

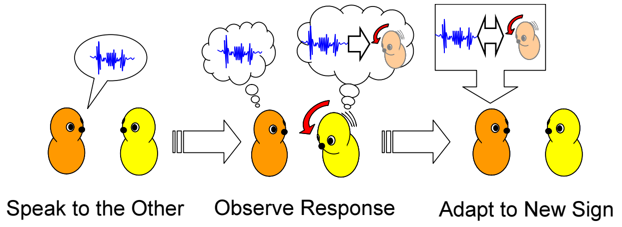

Fig. 11. Phase 2: Sharing of Sign Interpretation

IV. EXPERIMENTS

We carried out two sets of experiments. First, through imitation experiments, we evaluated the robots' capabilities to recognize signs generated by the other and to generate signs by themselves. Second, we carried out an interaction experiment, in which we observed the dynamic process of the interaction.

A. Experiment 1 - Evaluation of Capability for Recognition and Generation through Imitation

We trained Voice MTRNN and Motion MTRNN as described in Section III-B.1 and their configurations are listed in Table III. Time constants (τ) for them were both set for IO of 2, for Cf of 5, and for Cs of 70. In this experiment, Keepon **A** imitates the **B**'s actions as follows.

1) **B** makes an action.
2) **A** obtains the sensory information by observing the action, and translates it into $\boldsymbol{Cs_0}$ vector by MTRNN.
3) **A** makes an action generated by MTRNN from the $\boldsymbol{Cs_0}$ obtained in (2).

We evaluated the imitation error defined as the mean square error between the sequential data of motor information of **B**'s action in (1) and those of **A**'s action in (3).

The imitation errors in the voice and motion are listed in Table IV. The result of voice imitation whose imitation error is 0.00157 (closest to the average error of whole voice patterns) is plotted in Fig. 12. The result of motion imitation whose imitation error is 0.00162 (closest to the average error of whole motion patterns) is plotted in Fig. 13.

TABLE III

CONFIGURATION OF MTRNN FOR IMITATION EXPERIMENTS

	Voice MTRNN	Motion MTRNN
Cf Nodes	40	30
Cs Nodes	7	5
Input Data Steps	51 (30 msec/step)	45 (66.6 msec/step)
Training Patterns	110	70

TABLE IV

IMITATION ERROR

	Imitation Error	
	Voice	Motion
Average of Whole Patterns	0.00153 (124 patterns)	0.00154 (80 patterns)
Average of Training Patterns	0.00160 (110 patterns)	0.00157 (70 patterns)
Average of Validation Patterns	0.00101 (14 patterns)	0.00135 (10 patterns)

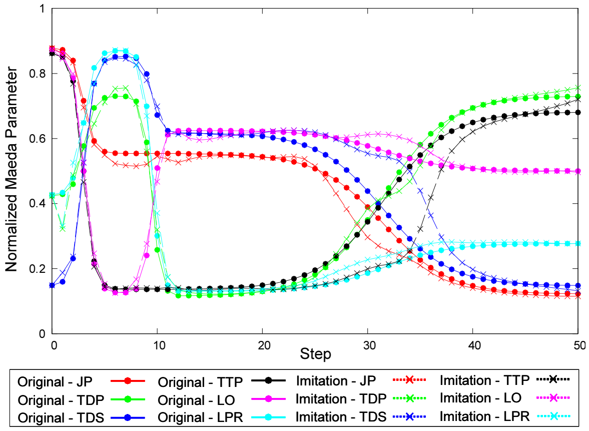

Fig. 12. Result of Voice Imitation (Imitation Error: 0.00157)

We confirmed from these results that the robots have capabilities of recognizing and generating signs by using the framework of sensori-motor integration. We also confirmed that they have the capability of generalizing their experience because they were able to imitate unknown validation patterns as well as training patterns (cf. Table IV).

B. Experiment 2 - Observation of Dynamic Process of the Interaction

We trained Voice MTRNN and Motion MTRNN as described in Section III-B.1, and their configurations are listed in Table V. Time constants (τ) for them were both set for IO of 2, for Cf of 5, and for Cs of 10000. We provided the two Keepons with the same Voice MTRNN and Motion MTRNN. Interpretation NN had 10 hidden layer nodes. The procedure described in Section III-B.2 was repeated 3000 times through computer simulation, and communication error was evaluated for each iteration. We defined **A**'s communication error as the mean square error between the sequential data of the motion that **A** primarily intended when speaking and those of the motion that **B** responded with (**B**'s communication error is

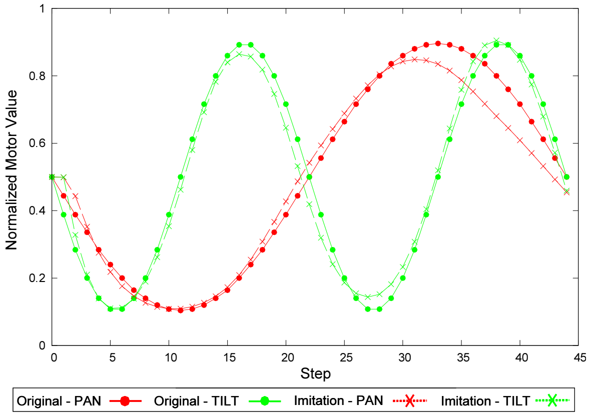

Fig. 13. Result of Motion Imitation (Imitation Error: 0.00162)

TABLE V

CONFIGURATION OF MTRNN FOR INTERACTION EXPERIMENT

	Voice MTRNN	Motion MTRNN
Cf Nodes	20	25
Cs Nodes	3	3
Input Data Steps	21 (50 msec/step)	20 (66.6 msec/step)
Training Patterns	12	20

similarly defined by interchanging the roles of **A** and **B**).

The results of the experiment are shown in Fig. 14. The graph at the top of Fig. 14 shows the sequence of communication errors for Keepons **A** and **B**. The others show the voice and motion patterns generated in segments I, II, and III of the interaction.

These results from the experiment revealed the following facts. There is repetition of coherent states with low error and incoherent states with high error in the interaction. In coherent states, both robots conveyed their intention to the other correctly, and interacted stably using similar voice-motion pairs (cf. segments I and III in Fig. 14). On the other hand, in incoherent states, they failed to convey their intention and exhibited irregular behaviors (cf. segment II in Fig. 14). The signs used for communication in coherent states (e.g., segments I and III) were different. Moreover, the voice and motion patterns used as signs differed from the training patterns described in the section III-B.1. The communication error tended to decrease on the whole, but the interaction kept evolving without convergence.

V. DISCUSSION

A. Evolutionary Process of Interaction

As a result of the experiment, we found that interaction with our model kept evolving through repeating coherent states, where the robots communicated with each other successfully, and incoherent states, where they miscommunicated. In the interaction, the robots created new diverse signs that depended on their body dynamics and experience through the framework of sensori-motor integration with MTRNN, and they shared the meanings of the signs through mutual adaptation. In conclusion, we confirmed the possibility of our approach to model evolutionary interaction in which diverse signs emerge and evolve through repetitive miscommunication and re-adaptation.

B. Modeling of Interaction Based on Dynamical Systems

Methods of modeling interaction or language that are based on multi-agent systems and on viewpoints that regard them as complex dynamical systems have attracted a great deal of attention [2]. Hashimoto utilized these methods in his study [3] as did Igari and Ikegami [4]. They also modeled interaction evolving through agents' mutual adaptation. Both of them dealt with the emergence of structures from various words or symbols in software simulations. On the other hand, we dealt with the emergence of signs themselves that depended on the agent robots' body dynamics and their experience.

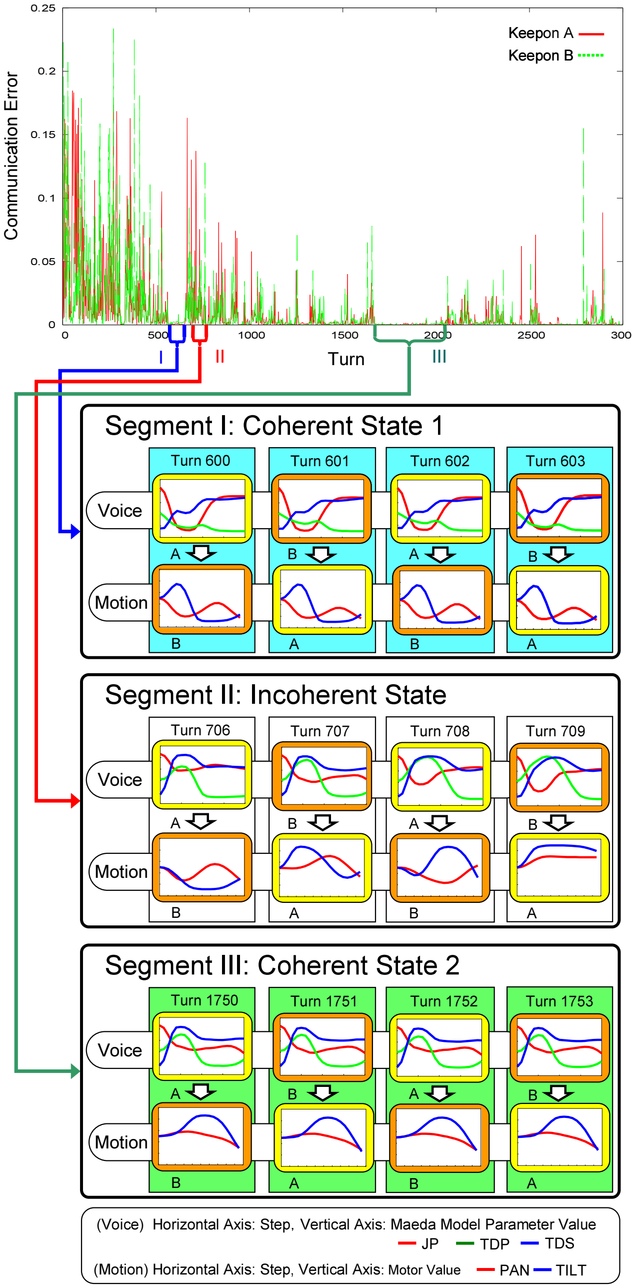

Fig. 14. Results of Interaction Experiment: Graph shows communication error sequence. Below graph are signs generated in Segments I, II, and III of interaction.

C. Interaction with Mirror System

The discovery of mirror neurons has revealed that humans (and primates) recognize others' actions reflecting their own body dynamics. For example, according to Liberman's motor theory of speech perception [10], humans recognize speech reflecting their own articulatory structure. This cognitive framework that matches observed events (sensory information) to similar, internally generated actions (motor information) is known as a mirror system. The mirror system forms a link between a observer and a actor, and is regarded as a necessary prerequisite for any type of communication [11]. Our sensori-motor integration framework by MTRNN is a model of the mirror system.

D. Dynamical Systems with Semiotics

Semiotics [12] is the study of signs and of their meaning and use. Charles S. Peirce, a founder of semiotics, offered a triadic model of a sign composed of a "sign" (to be exact, a "representamen"), an "interpretant" and an "object". In his model, a subjective dynamic process called "semiosis" where a sign gets connected to an object through an interpretant is emphasized. This is in contrast to an objective static structure of a connection between a sign and an object. Umberto Eco expanded the term "semiosis" offered by Peirce to designate a social dynamic process where society produces signs and attributes shared meanings to these signs. The process in our model, where a voice/motion sign becomes connected to a motion/voice through an Interpretation NN proper to each agent, corresponds to the process of semiosis offered by Peirce. The process of sharing signs by mutual adaptation through the interaction in our model corresponds to the social semiosis expanded by Eco. Semiosis has attracted much attention recently from researchers in engineering fields. There is even a project called "Design Theory for Dynamical Systems with Semiosis" [13].

E. Symbolic Interactionism

Symbolic interactionism is a sociological, social-psychological perspective based on following three premises [14].

1) Human beings act toward things on the basis of the meanings that the things have for them.
2) The meanings are derived from or arise out of the social interaction that one has with one's fellows.
3) These meanings are handled in, and modified through, an interpretive process used by a person in dealing with the things he/she encounters.

Our interaction system is a model of symbolic interactionism. The three premises in our model correspond to the following.

1) Keepons act (move/speech) in response to the other's signs (voice/motion patterns) on the basis of the meanings that the signs have for them.
2) The meanings are derived from or arise out of the interaction between Keepons.
3) These meanings are handled in, and modified through, an interpretation NN used by a Keepon in dealing with the signs it encounters.

From this perspective, the meanings of things (signs) in the society continuously, dynamically change and evolve through repeated formation and re-formation without given specific goals. Our experiments produced results that were consistent with this.

VI. Conclusions

We proposed an interaction model from the point of view that the bottom-up emergence of signs through agents' mutual adaptation is essential for evolutionary interaction. As a result of our experiments on imitation, we confirmed from our model that the agent robots could self-organize their cognitive structures through generalizing experience, and that they had the ability to recognize each other's actions in the real world. We carried out an interaction experiment and observed its dynamic evolutionary process. As a result, we confirmed that the interaction in our multi-agent system kept evolving through repeating alternately coherent states where agents successfully communicated with each other and incoherent states where they miscommunicated. In such interaction, diverse new signs (motion and voice patterns) depending on the robots' body dynamics emerged through the generalization capability of MTRNN. In conclusion, we found that our approach can be used to model evolutionary interaction in which diverse signs emerge through repetition of miscommunication and re-adaptation.

In future work, we intend to introduce contextual information to Interpretation NN to deal with the emergence of language that has syntax. We also aim to improve our interaction model so that it becomes a system in which turn-taking phenomena emerge.

VII. Acknowledgments

This research was partially supported by a Grant-in-Aid for Scientific Research, (S) 19100003, Creative Scientific Research, 19GS0208, Global COE, and RIKEN.

References

[1] Y. Miwa, S. Wesugi, C. Ishibiki, and S. Itai, "Embodied interface for emergence and co-share of 'Ba', usability evaluation and interface design," *Proc. HCI Int.*, pp. 248-252, 2001.

[2] L. Steels, "The synthetic modeling of language origins," *Evolution of Communication*, 1997.

[3] T. Hashimoto, "The Constructive Approach to the Dynamic View of Language," *Simulating the Evolution of Language*, 2001.

[4] I. Igari and T. Ikegami, "Coevolution of Mind and Language," *int'l conf. on Evolution of Language*, 2000.

[5] Y. Yamashita and J. Tani, "Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: a Humanoid Robot Experiment," *PLoS Comput. Biol.*, vol. 4, 2008.

[6] J. Tani and M. Ito, "Self-Organization of Behavioral Primitives as Multiple Attractor Dynamics: A Robot Experiment," *IEEE Trans. on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 33, no. 4, pp. 481-488, 2003.

[7] H. Kozima, C. Nakagawa, and H. Yano, "Using robots for the study of human social development," *AAAI Spring Symposium on Developmental Robotics*, 2005.

[8] S. Maeda, "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," *Speech production and speech modelling*, Kluwer Academic Publishers, pp. 131-149, 1990.

[9] H. Kanda, T. Ogata, K. Komatani, and H. G. Okuno, "Segmenting Acoustic Signal with Articulatory Movement using Recurrent Neural Network for Phoneme Acquisition," *IEEE/RSJ IROS2008*, 2008.

[10] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, 1985.

[11] G. Rizzolatti and M. A. Arbib, "Language within our grasp," *Trends in Neurosciences*, pp. 188-194, vol. 21, 1998.

[12] D. Chandler, "Semiotics for Beginners," online, 1995,
http://www.aber.ac.uk/media/Documents/S4B/semiotic.html.

[13] T. Sawaragi, et al. , "Design Theory for Dynamical Systems with Semiosis," online,
http://www.syn.me.kyoto-u.ac.jp/semiosis/en/index.html.

[14] H. Blumer, "Symbolic interactionism: perspective and method," University of California Press, 1986.