Paper:

Constructive Approach to Role-Reversal Imitation Through Unsegmented Interactions

Tadahiro Taniguchi[*], **Naoto Iwahashi**[**,***], **Komei Sugiura**[***], **and Tetsuo Sawaragi**[****]

[*]Department of Human & Computer Intelligence, Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan
[**]National Institute of Information and Communications Technology
4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan
[***]Advanced Telecommunications Research Institute International
2-2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-0288, Japan
[****]Graduate School of Engineering, Kyoto University
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan
E-mail: tadahiro@tanichu.com

[Received February 6, 2008; accepted June 12, 2008]

This paper presents a novel method of a robot learning through imitation to acquire a user's key motions automatically. The learning architecture mainly consists of three learning modules: a switching autoregressive model (SARM), a keyword extractor without a dictionary, and a keyword selection filter that references to the tutor's reactions. Most previous research on imitation learning by autonomous robots targeted motions given to robots, were segmented into meaningful parts by the users or researchers in advance. To imitate certain behavior from continuous human motion, however, robots must find segments to be learned. To achieve this goal, the learning architecture converts a continuous time series into a discrete time series of letters using the SARM, finds meaningful segments using the keyword extractor without a dictionary, and removes less s meaningful segments from keywords using the user's reactions. In experiments, an operator showed unsegmented motions to a robot, and reacted to the motions the robot had acquired. Results showed that this framework enabled the robot to obtain several meaningful motions that the operator hoped it would acquire.

Keywords: imitation learning, self-organized learning, role-reversal imitation, switching linear model, keyword extraction

1. Introduction

If we are to develop autonomous robots that can live together with us in society, they must be able to acquire various concepts and behavior naturally while they spend their time with us. However, there are still very few methods of giving robots such capabilities. If we had to teach all behaviors to autonomous robots when they interacted with us in our daily lives at some point in the future, it would be too time-consuming to do so. One of the first attempts at user-robot coexistence involved robot pets such as Sony's Aibo. However, most of these pets did not change the way they interacted with their owners, and their behavior was limited. Consumers were noticed to become bored interacting with these unchanging vehicles of entertainment. The limitations in their adaptability made their owners disinterested. In contrast, living pets, like dogs continue to fascinate us. This tells us that the appearance of robots and their designed behaviors are important in terms of short-term interactions. However, adaptability and the ability to develop are more important for entertainment robots. If an entertainment robot, which can learn different behaviors incrementally and autonomously without any explicit commands or signals, is developed, it will have a great impact on society. If users must give explicit commands, they feel that robots have not acquired behavior heteronomously rather than autonomously. This prevents users from empathizing with robots. This emotional distance must be reduced for people to accept robots as quasi-human entities. However, such emotional aspects are also crucial in the field of entertainment robots. Therefore, autonomous robots are expected to acquire motions incrementally by themselves. In contrast, human children usually acquire a large number of motions and gestures through interactions with their parents and other people through imitation-learning strategies. Here, we have especially focused on "role-reversal imitation". In many types of imitation-learning strategies, role-reversal imitation is believed to be typical of the human race. It is important to construct a computational model for role-reversal imitation through unsegmented interaction both to develop robots that learn adaptively and to understand the human capabilities of imitation.

In this paper, we propose a novel method of machine learning that enables robots to acquire multiple behaviors incrementally through natural, continuous human-robot interaction.

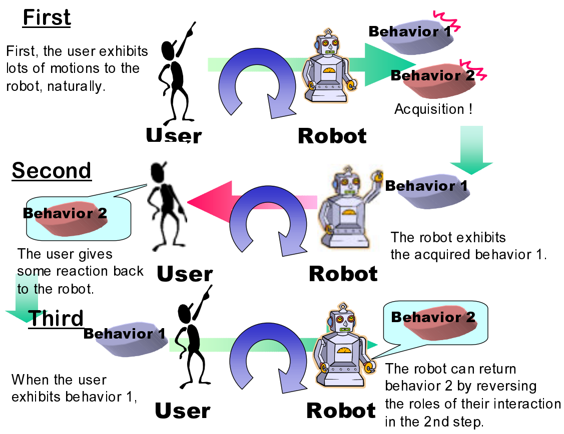


Fig. 1. Role-reversal imitation.

1.1. Role-Reversal Imitation Learning from Unsegmented Interaction

Imitation learning is a very powerful and natural learning strategy for human beings. This will also enable a robot to learn new motions much faster than other schemes such as reinforcement learning because a robot can directly utilize the trajectories of motions of others.

However, it is difficult to identify segments that a robot should learn because a user's motion directed toward the robot is an unsegmented continuous time series. In most previous research on imitation learning by autonomous robots, target motions that robots were made to learn were segmented into meaningful parts by operators beforehand [2, 6, 12]. To overcome this problem, it is important for the robots themselves to extract meaningful segments from the natural continuous human motions.

Tomasello, in studying how children imitate adults [16], he asserted that "role-reversal imitation" is an important concept in research on imitation learning. When a child demonstrates motion A to his or her mother, she will respond with some motions in return. Here, we have assumed the mother's motion is motion B. By repeating such interactions, the child imitates his or her mother's behavior and acquires motion B. Additionally, in role-reversal imitation, the child imitates not only the behavior itself (motion B), but also his or her mother's role in the interaction. Once the child has acquired the role of his or her mother, he or she responds with motion B in return when his or her mother demonstrates motion A. Thus, the roles of the mother and the child are now reversed, with the child playing the role of the tutor. This is role-reversal imitation (see **Fig. 1**).

In this natural framework, we should be able to develop a robot that can acquire multiple behaviors and ways of using them through natural human-robot interactions.

However, it is also difficult to identify the segment that is the mother's reaction to the child's action because her motion is a continuous time series. To overcome this problem, it is again important for the robot to extract meaningful segments from a human's natural flow of continuous motion. If the robot detects the key motion from the human's continuous flow, it can easily correlate its own motion with the human's key motion. The contingency[1] of the human's reaction will confirm the meaningfulness of the robot's motion.

To achieve role-reversal imitation through unsegmented interactions, we utilized the switching autoregressive model (SARM) [10] to segment a continuous time series and determine autoregressive (AR) models that represented all the segmented dynamics. However, these obtained segments are not usually meaningful. The SARM segmented the time series by simply referring to its dynamical properties, in other words, all segmented dynamics are linear. Most of the time, segments seem to be meaningless to users. However, certain sequences of segmented motions are meaningful. If we assume that the segmented motions are represented by letters corresponding to hidden states of the SARM, the time series are converted into "*documents*," which are sequences of letters. We considered the meaningful sequences of segmented motions to have similar properties to keywords in documents written in natural languages such as Japanese, English, and Chinese. In other words, we assumed that n-grams, which are converted from the key motions of the time series presented by the robot's user, are distributed over the set of time series similarly to the keywords of documents written in natural languages [2].

We utilized the keyword extractor proposed by Umemura to find keywords in the documents [17].

However, the group of keywords obtained by utilizing the above two methods of segmentation and keyword extraction still includes a few motions that are meaningless to users. To remove such meaningless motions from keywords, our learning architecture utilized the reaction of a user to the motions demonstrated by the robot using acquired keywords. We assumed that the users would usually return the robot's meaningful motions with a meaningful response. Therefore, the robot "considered" motions that often induced users to exhibit meaningful motions to be real key motions. A keyword-selection filter removed meaningless keywords from those obtained by referring to the user's reactions.

Based on these assumptions, our method enabled a robot to extract a user's meaningful motions, i.e., key motions.

1.2. Related Work

The idea of the emergence of behavior has been attracting attention [1], which enables robots to evolve through interactions with their environment and/or human users. In particular, if an entertainment robot acquires several behaviors through interactions with its owner, that will make the user more likely to have greater empathy with

1. Contingency, a developmental psychology term [18], refers to a sequential, turn-taking dyadic structure assumed to be an optimal form of social stimulation for infants. It is widely believed that certain optimal infant-caregiver social structures facilitate a child's social, emotional, and cognitive development [5, 9].
2. Whether this assumption is valid should be studied in further cognitive research to justify our approach.

the robot. Therefore, the emergence of behavior should be studied from the viewpoint of entertainment robotics.

Taniguchi et al. proposed a reinforcement learning schema model (RLSM), which enables a robot to acquire several behaviors based on the framework of modular reinforcement learning [15]. However, reinforcement learning usually requires a huge amount of time to learn a new behavior. Therefore, it is almost impossible for a robot to learn several behaviors using RLSM through natural human-robot interactions because learning several behaviors based on reinforcement learning would take a very long time.

Imitation learning is an effective framework for making robots acquire various motions. Basically, imitation learning is not based on trial and error, but on supervised learning. Therefore, it can be applied to humanoid robots that have many degrees of freedom (DOFs). It is extremely difficult for reinforcement learning to cope with such high-dimensional systems.

Inamura et al. [6] and Sugiura et al. [13] made robots acquire several behaviors by using a continuous hidden Markov model (HMM). Inamura et al. called this the mimesis model. However, in their research, the behaviors that the robot had to learn were completely separated before the robot started to learn, e.g., to "walk," "squat," and "kick." Therefore, they did not treat how imitators discriminated what they should imitate from a continuous time series of interactions. Sugiura et al. did not address this problem either.

Ito et al. made a small humanoid robot, QRIO, to learn several behaviors from a user's continuous motions. In one experiment, a human experimenter grasped QRIO's arm and moved it. From this display of target motions, QRIO acquired several behaviors by using a recurrent neural network with parametric bias (RNNPB) [7]. Yokoya also utilized RNNPB to make a robot learn how to move an object on a table by imitation and to obtain a generalized concept of this motion [19]. Although RNNPB is a single recurrent neural network, it can be used to obtain several motion patterns because it has several parametric biases that augment the RNN's parameter space. However, when RNNPB learns several behaviors incrementally, it is important for the experimenter to demonstrate the same motion pattern for a long period enough for it to be learned. Their model usually also expected several motion patterns to successively switch from one meaningful pattern to another that was also meaningful. However, human children acquire several behaviors by observing their parents' natural daily motions including meaningful motions worth imitating together with meaningless motions demonstrated to children unconsciously.

Whatever the key factor may be that determines when and what to imitate, a robot will be able to learn behaviors through natural human-robot interactions if it can extract meaningful segments from natural continuous motions demonstrated by humans.

From another perspective, many researchers have been studying the extraction of various characteristics from time-series data, e.g., the extraction of "motifs" or "motion primitives" [3, 8]. These motifs are previously unknown patterns that appear frequently in time-series data. Tanaka et al. proposed an algorithm that could discover time-series motifs from multi-dimensional human motion data based on the minimum description length (MDL) principle [14]. However, most of these methods are usually only used to extract "motifs" or "motion primitives." To achieve imitation learning from unsegmented time series, the algorithm must not only extract motion primitives, but also generate the acquired motion. Due to this reason, recurrent neural networks (RNNs) [7], HMMs [6, 13] or polynomial equations [12] have been utilized to express motion primitives in research on imitation learning. These are similar to our research. However, the "motifs" of a mother's behaviors are not necessarily a time series, which a child should imitate.

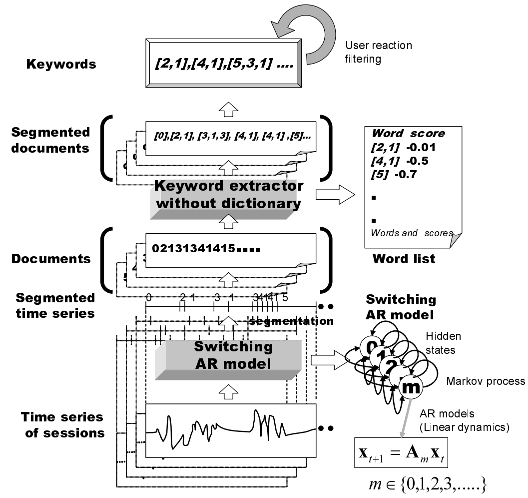


Fig. 2. Overview of our model.

Based on these related studies, we devised a method of imitation learning. It is a type of self-organizing learning architecture that enables a robot to segment the continuous motions of human users to extract key motions from them, recognize meaningful reactions exhibited by human users, and choose really meaningful key motions from the obtained key motions with respect to the human users' reactions. The total learning architecture achieved a self-organizing learning through continuous human-robot interactions. This learning method cannot be categorized as either supervised or reinforcement learning. The learning process can be considered as a self-organized or unsupervised learning process.

2. Computational Role-Reversal Imitation Learning

2.1. Overview

There is an overview of our framework for extracting key motions from continuous motion data in **Fig. 2**. There are two major components in this model. The first is the SARM proposed by Murphy [10], and the second is keyword extraction without dictionaries (section 2.3) proposed by Umemura [17]. A session starts when a user enters a room where there is a robot, and it ends when he or she exits the room. The robot records the user's motion while he or she interacts with it and obtains a "time series of a session" (**Fig. 2**). We assume that we could obtain several time series by recording "sessions" of human-robot interactions. We also assume that each session has its own topic to some extent. SARM could decompose a continuous time series into several linear autoregression models [3]. Therefore, the time series can be translated into a sequence of indices of SARM's hidden states. We called the sequences "documents" and denoted each hidden state by a letter. After the time series was translated into documents, the documents were segmented and several keywords embedded in the documents written in the letters were extracted by using Umemura's method of text mining [17].

Most text mining methods utilize a prepared dictionary to extract keywords from documents. However, Umemura's does not utilize any dictionaries. He demonstrated that keywords could be extracted without any dictionaries if there were enough Japanese documents and the documents satisfied the assumption that there was a disproportional distribution of keywords. We assumed that the documents obtained by applying SARM to the sequences of human-robot interactions also satisfied this condition. By considering sequences of hidden states as documents in an unknown language, we applied Umemura's method to extracting keywords that represented key motions of a human user from the documents. In our model, the extracted keywords seemed to correspond to key motions embedded in the time series that the user displayed. However, the extracted key motions did not always satisfy the user's standard of "what is a meaningful motion?". We assumed that users would return some meaningful motions in turn when they observed the robot exhibiting some of these. By utilizing this assumption, the robot could erase keywords that were not very important from its keyword list.

2.2. Switching Autoregression Model (SARM)

SARM models a multidimensional time series by using multiple AR models whose hidden states are switched from one to another based on a Markov process. SARM has M hidden states. We denote the hidden state at t by s_t. A hidden state j has an AR model.

$$x_t = A_j x_{t-1} + v_t \quad \ldots\ldots\ldots\ldots (1)$$

where j represents the index of the hidden states, x_t represents a state variable at t, and $v_t \sim N(0, Q_t)$. First, we will explain how to determine which AR model should be chosen under the condition where the robot observes a time series, $x_{1:T}$. To achieve that purpose, the probability, $\Pr(s_t = j|x_{1:T})$, should be calculated. Suppose that the hidden state has Markov properties,

$$\Pr(s_t = j|x_t, x_{1:t-1}) \quad \ldots\ldots\ldots\ldots (2)$$

$$= \frac{1}{c}\Pr(x_t|s_t = j, x_{1:t-1})\Pr(s_t = j|x_{1:t-1}) \quad \ldots\ldots (3)$$

$$= \frac{1}{c}\Pr(x_t|s_t = j, x_{1:t-1})\sum_i \Pr(s_t = j|s_{t-1} = i, x_{1:t-1})$$

$$\times \Pr(s_{t-1} = i|x_{1:t-1}) \quad \ldots\ldots\ldots\ldots (4)$$

$$= \frac{1}{c}L_t(j)\sum_i Z(i,j)\Pr(s_{t-1} = i|x_{1:t-1}) \quad \ldots\ldots (5)$$

where c is the normalization constant and Z is the transition matrix of hidden states. $Z(i,j)$ represents the probability that the hidden state will transit from i to j in one time step.

$$L_t(j) = N(x_t; A_j x_{t-1}, Q_j) \quad \ldots\ldots\ldots (6)$$

is the likelihood of prediction error at time t given by the AR model, j, where N is a multidimentional normal distribution whose center vector is $A_j x_{t-1}$ and whose variance-covariance matrix is Q_j. On the backward pass, we have

$$\Pr(s_t = j|x_{1:T}) = \sum_k \frac{\Pr(s_t = j|x_{1:t})\Pr(s_{t+1} = k|s_t = j)}{\Pr(s_{t+1} = k|x_{1:t})}$$

$$\times \Pr(s_{t+1} = k|x_{1:T}). \quad \ldots\ldots\ldots (7)$$

The derivation of these equations is almost the same as the derivation of the HMM. The parameters, A_j, Q_j, Z, can be estimated using the following expectation maximization (EM) algorithm below [10][4]:

$$A_j = \left(\sum_l \sum_{t=2}^{T} W_t^j P_{t,t-1}\right)\left(\sum_l \sum_{t=2}^{T} W_t^j P_{t,t-1}\right)^{-1} \quad \ldots (8)$$

$$Q_j = \left(\frac{1}{(\sum_l \sum_{t=2}^{T} W_t^j)}\right) \times$$

$$\left(\sum_l \sum_{t=2}^{T} W_t^j P_t - A_j \sum_l \sum_{t=2}^{T} W_t^j P'_{t,t-1}\right) \quad \ldots (9)$$

$$Z(i,j) = \frac{\sum_l \sum_{t=2}^{T} \Pr(s_{t-1} = i, s_t = j|x_{1:T})}{(\sum_l \sum_{t=2}^{T} W_t^j)} \quad \ldots (10)$$

where $W_t^j \equiv \Pr(s_t = j|x_{1:T})$, $P_t \equiv x_t x_t'$, and $P_t \equiv x_t x_{t-1}'$. In the original paper [10], the initial distribution, $P(s_1 = j) = \pi_j$, was also re-estimated. However, we have not re-estimated this parameter for the sake of simplicity. We have not updated $Z(i,j)$ either.

Next, we extract a "document" from the time series of the posterior probabilities.

$$s_t^* = \underset{j}{\mathrm{argmax}}\Pr(s_t = j|x_{1:T}) = \underset{j}{\mathrm{argmax}}\, W_t^j \quad \ldots (11)$$

where s_t^* is a sequence of the most likely hidden states. By ignoring identical neighboring states, we compress sequences containing the same successive letters into a document that does not contain any repetitive letters (see

3. SARM can be considered to be a type of HMM.

4. This is also called the Baum-Welch algorithm.

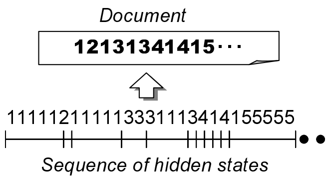


Fig. 3. Compression from time series of hidden states to document.

Fig. 3). Finally, we obtain the "documents". The number of documents is the same as the number of sessions in which a user interacts with the robot. After that, the robot tries to find keywords from the documents that may appear meaningless at first sight. Generally, each n-gram of hidden states of the SARM, e.g., [1, 3, 4, 5], and [2,1,2,1], represents a segment of motion. However, most motions generated by these n-grams seem to be meaningless to the users. For example, if the SARM has 10 hidden states and motions are generated by 4-grams, the number of candidates of key motions become about 7000 - too large for a user to teach the robot meaningful motions through interaction by reacting to the robot's behavior generated by using these candidates. Keyword extraction is thus needed to find keywords from documents. Therefore, making powerful keyword extraction method is crucial.

2.3. Keyword Extraction Without Dictionaries

Umemura proposed a useful method of extracting keywords that did not require the use of a dictionary [17]. Unlike in English, it is difficult in Japanese to extract meaningful segments from documents without using dictionaries within the context of research on extracting keywords or terms from documents because Japanese, unlike English, does not have explicit segmentation. In much previous work on keyword extraction, researchers utilized morphological-analysis approaches using various dictionaries to segment target documents unsegmented in Japanese.

Umemura assumed that keywords would appear disproportionately in documents. If a document's topic is relevant to a keyword, that keyword will appear several times in the document. This interesting quantitative tendency has been found in documents written in natural language by Church [4]. The keyword, on the other hand, rarely appears in other documents. Based on this assumption, Umemura introduced some very simple criteria for calculating the scores of n-grams[5]. The higher the score, the more likely the n-gram is to be a keyword.

The score is calculated based on a statistic termed "positive adaptation" [4]. Church defined positive adaptation as

$$Pr(+adapt) = Pr(k \geq 2 | k \geq 1) \approx DF_2/DF_1 \quad . \ (12)$$

where DF_k (document frequency k) is the number of documents that contains the n-gram k or more times. DF_k is the generalization of document frequency DF, a standard term in information retrieval. DF is the number of documents that contains the target n-gram.

However, positive adaptation is insufficient for identifying keywords from documents because the positive adaptation of a keyword's substrings often has a similar value to that of the keyword itself. For example, "watermelo," which is a substring of "watermelon," appears in all documents almost in the same way as "watermelon" itself.

To overcome this problem, Umemura's method segments the documents based on the score, which represents how much the n-gram seems to be a single meaningful word. That segments the documents to maximize the sum of their scores. The boundaries of n-grams are determined by this rule.

The score of an n-gram is calculated by

$$score = \log_2(\min(\mathrm{UB}, Pr(+adapt))) \quad . \ . \ . \ . \ (13)$$

where UB (upper bound) is a metaparameter that can be determined by a designer. Generally, when UB approaches 1.0, the target document will be finely segmented. In contrast, when UB approaches 0.0, the target document will be coarsely segmented. If $DF_2 > \mathrm{MA}$, the n-gram is placed in the "wordlist" with its score (see **Fig. 2**). The MA (minimum appearance) is a metaparameter that determines the size of the wordlist. In Umemura's research, the threshold values, i.e., UB and MA, were set heuristically. The experimental results depend on these values.

If the target n-gram is randomly distributed over all documents, $Pr(+adapt) = DF_1/DN$, where DN (document number) means the total number of documents. However, it is well known that the second instance of a word (or n-gram) is much more likely than the first to appear in documents written in natural language because of the lexical content. Therefore, we simply determine if the n-gram is the keyword of the topic of some documents by observing $Pr(+adapt)$. Additionally, *log* is only applied to scaling.

To find the best segmentation of a document, a Viterbi search is utilized to reduce the computational effort involved. During this computation, the scores of n-grams that are not registered in the word list are assumed to be -10000.

After segmentation, the keyword extractor extracts keywords from the segmented documents. In our experiment, n-grams in segmented documents that satisfied the following conditions are considered to be keywords.

1. $DF_2/DN < F_{max}$
2. $score > score_{min}$
3. A keyword contains more than one letter.

This method of extracting keywords has four metaparameters, i.e., UB, MA, F_{max}, and $score_{min}$. At the moment, they have to be determined heuristically. Our future work is to investigate how to determine these values based on statistical theory.

5. n-gram means successive n letters in a document

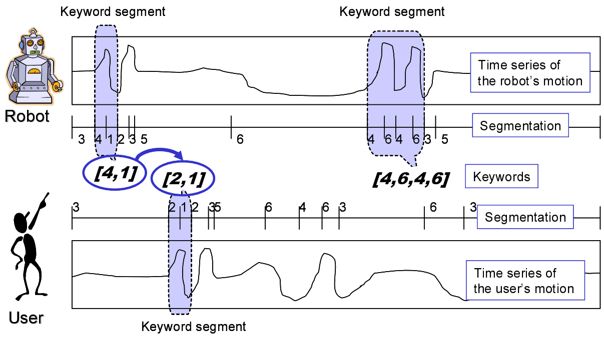


Fig. 4. Robot can discriminate user's reactions by utilizing acquired SARM, word list, and keywords.

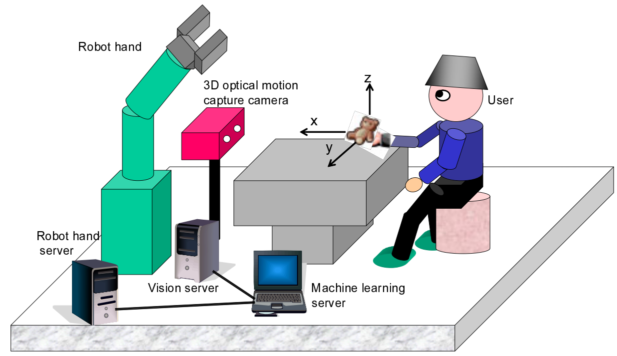


Fig. 5. Overview of experimental environment.

2.4. Recognition of Human Users' Reactions Based on Extracted Keywords

After the robot has acquired (1) SARM, (2) the word list, and (3) keywords, it can use them and exhibit keymotions. Once the robot has selected a keyword, AR models are selected hidden states in order corresponding to the characters contained in the keyword and the robot starts to move [6].

When a robot exhibited an interactively meaningful motion, e.g., waving good-bye or saluting, we assumed that the user in front of it would respond with some interactively meaningful motion. The robot could recognize the user's reaction as a keyword by utilizing the obtained SARM and a keyword list if the reaction was meaningful (**Fig. 4**). Therefore, we made the robot observe what the user did after it had demonstrated a key motion. If the user's motion was recognized as a keyword, the robot could detect that its own motion was meaningful to the user (positive feedback). By utilizing this relation, the robot could remove meaningless keywords from those that it had obtained, i.e. negative feedback.

As previously mentioned, the robot could interpret the user's reactions by utilizing the acquired SARM, word list, and keywords. If the user usually demonstrated a series of motions, which was recognized as a keyword, back to the robot after it had demonstrated its motions, the keyword could be considered as a response to the keyword that produced the robot's motions. By observing these pairs of actions and reactions, the robot could learn its response to the user when he or she demonstrated the key motion. Therefore, our model enables robots to recognize human users' reactions from continuous human-robot interactions.

This filtering strategy utilizing a user's reaction is similar to reinforcement learning. Some readers might think that our approach is not sufficiently efficient or direct. Giving a pre-defined reinforcement signal to the robot is the simplest approach to filtering non-meaningful motions. However, our approach is superior to this simplest approach in two respects. The first is that the human user can interact with the robot without any explicit rules given by the designer. The strategy adopted in our learning method is totally different from that of reinforcement learning, where reinforcement signals, in other words, rewards and penalties, are determined a priori. However, the user's reactions in our model, which give positive feedback to the robot, are also learned through continuous human-robot interactions. Our learning architecture does not employ either teaching signals that gives the learning architecture what the key motions are, or predefined reinforcement signals that gives the learning architecture an information which motions are more likely to be meaningful. This means that our imitation-learning framework is a self-organizing method of learning. Therefore, this architecture could possibly enrich human-robot interactions, in terms of adaptability. Secondly, our approach is superior because it is more reasonable as a constructive model of role-reversal imitation by human children. In daily life, a child autonomously imitates his or her mother's motions. If a parent is paying attention to the child and responds with some reactions to him or her, he or she is usually pleased and increases the frequency of the motion. In psychological terms, children learn the social meanings of their motions based on such contingencies. In such daily learning processes by children, mothers rarely directly punish or inflict penalties on them. Therefore, our approach is more reasonable from the viewpoint of being a constructive scheme to children's imitative learning.

3. Experiment

To test our framework, we carried out an experiment in a human-robot-interaction environment.

3.1. Conditions

We prepared a robot that had a long hand, and a 3D motion-capture camera to record human-robot interactions in the real world (**Fig. 5**). The x-, y- and z-axes values of the experimenter's hand positions could be measured with the 3D motion camera. Each value at time t was represented by $x_t^o, y_t^o, and z_t^o$. We deal only discusses

6. After keywords have been extracted, the average staying time for all hidden states included in each keyword is calculated from the original time series. The average staying time for all hidden states in a keyword is obtained in a corresponding time series. The initial state of all key motions is also calculated in the same way. By utilizing this information, the robot can decode keywords written in letters into key motions in the real physical world.

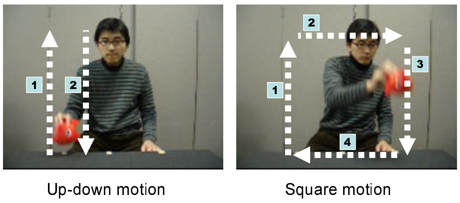


Fig. 6. Two motions that experimenter embedded in time series.

the imitation problem for hand positions for the sake of simplicity. In other words, we have ignored elbow and shoulder positions. To model a multi-dimensional time series was not our central issue here because many researchers have already treated this problem [6, 12]. The robot was operated by referring to target positions. If the robot server output a target position to the robot, the robot's hand position moved to that position. When the target positions were successively given or outlier positions were given, the robot smoothed the given time series and moved based on the smoothed time series. The robot could observe the experimenter's hand position, and observable state variables were defined to be $x_t = (x^o_t, y^o_t, z^o_t, c)$ where $c = 100$ is a constant term and the each axis are in millimeters $[mm]$. We assumed that the imitator could observe the tutor's posture and directly map it to its own posture by using an identical transformation. The "correspondence problem" [11] is considered to be a serious issue in research on imitation learning. It suffers from several elemental difficulties. One of these is "how can an imitator match the performer's state variables to his/her own state variables?", i.e., transform the coordinates. We ignored this problem in this paper. We gave the true transformation matrix to the robot. Therefore, the robot could map the experimenter's hand position to that of the robot's hand.

In this experiment, the experimenter displayed a series of motions to the robot in each of twenty-one sessions. One session lasted for about twenty seconds[7]. The sampling rate was 25 Hz. In each session, the experimenter exhibited movements that were characteristic of that session. We defined two sorts of target motions, as shown in **Fig. 6**.

In sessions 1-7, the experimenter demonstrated the up-down motion several times to the robot. In sessions 8-14, he demonstrated the square motion several times to it. In the other sessions, the experimenter randomly displayed the up-down motion, the square motion, and some other meaningless motions. In these sessions, the experimenter did not explicitly segment the time series. Therefore, the robot could not explicitly determine which region of the time series was the up-down motion or the square motion. The robot also did not know how many sorts of motions the experimenter intended to teach it. There are examples of unsegmented motion demonstrated by the experimenter in **Fig. 7**.

In this experiment, meaningful motions were not necessarily repeated, and meaningless motions were randomly inserted. Therefore, extracting a meaningful segment from a continuous human motion was not an easy task. There is an example of a time series and its hidden states estimated by SARM in **Fig. 8**.

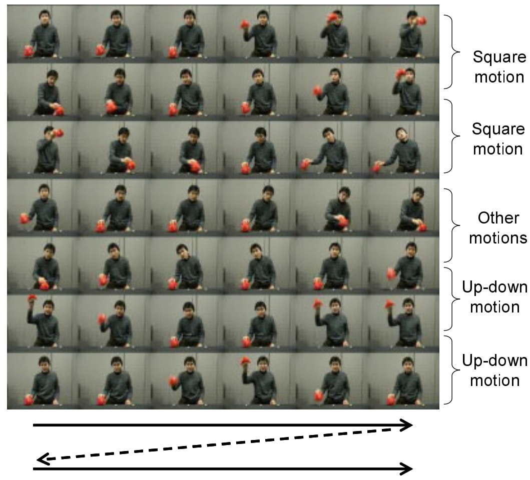


Fig. 7. Example of unsegmented time series of human-hand motion.

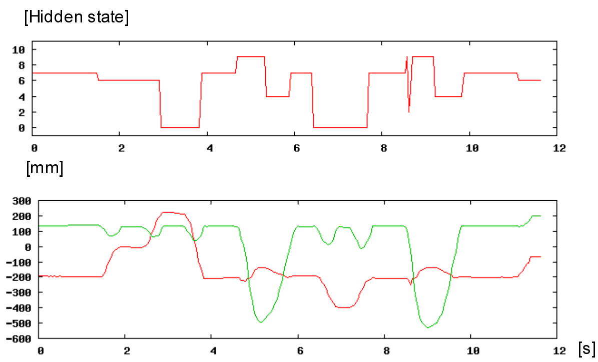


Fig. 8. Top: hidden states estimated by SARM, bottom: a time series exhibited to the robot by an experimenter.

After this, SARM parameters $(\{A_j, Q_j\})$ were estimated by utilizing the time series of the twenty-one sessions. We kept the transition matrix, Z, constant during this experiment. The diagonal elements of Z were set to 0.964, and the others were set to 0.004. We defined ten hidden states heuristically. The initial parameters of the AR model were given at random. The EM algorithm was iterated five times to estimate the parameters. By utilizing the estimated parameters, SARM computed the posterior probabilities, $\Pr(s_t = j|x_{1:T})$, for each session (see Eq. (7)). The sequences of most likely hidden states $\{s^*_t\}$ were easily determined from the posterior probabilities (see Eq. (11)). By compressing the sequences, the robot obtained twenty-one "documents." Keyword extraction was then applied and finally the robot acquired several keywords and a word list. In this experiment, the keyword-extractor's parameters were set to $\{UB = 0.9, MA = 2, F_{max} = 0.4, score_{min} = -0.5\}$.

7. How long each session lasted was left to the experimenter's discretion. As a result, each session lasted about twenty seconds.

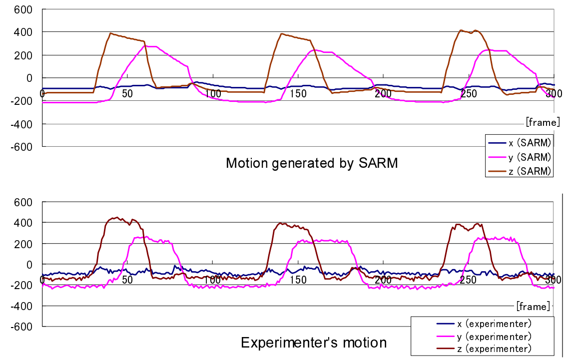


Fig. 9. Top: Time series generated by robot by utilizing estimated SARM and estimated most-likely hidden states in tenth session. Bottom: Original time series conducted by experimenter in tenth session.

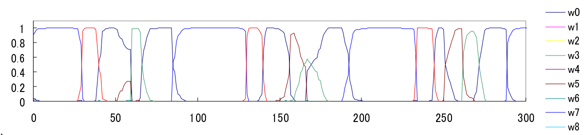


Fig. 10. Transitions in posterior probability W_t^j for all AR models.

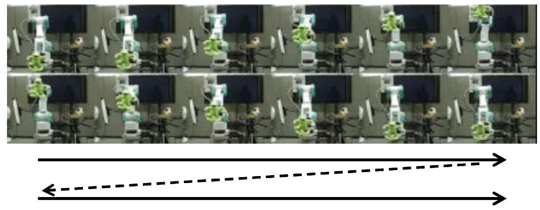


Fig. 11. Trajectories produced by robot utilizing acquired keywords [7,9,4,3] corresponding to up-down motion.

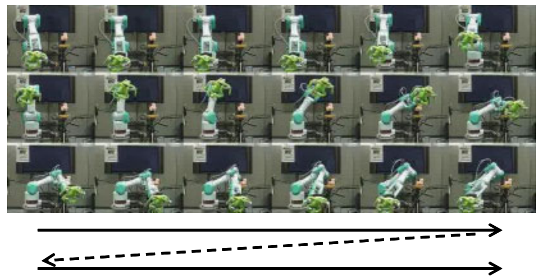


Fig. 12. Motion produced by robot utilizing acquired keywords [7,9,0,5,3,0,7] corresponding to square motion.

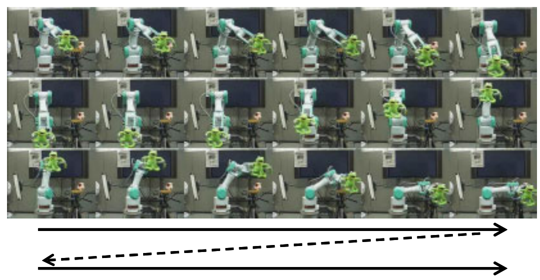


Fig. 13. Motion produced by robot utilizing acquired keywords [0,7,9,0,5] that were not intended for acquisition.

3.2. Results

SARM parameters were estimated, and the word list was also calculated. The time series produced by the estimated SARM and the estimated most-likely sequence of hidden states in the tenth session are shown in **Fig. 9** as an example. We can see that SARM, which contained ten AR models, could model the experimenter's unsegmented motion to some extent. The transition in posterior probabilities W_t^j that represents which linear model seemed adequate at each instant in time is shown in **Fig. 10**.

After the keyword extractor was utilized, seven keywords were extracted. They were $[7,9,0,5,3,0,7]$, $[9,0,5,3,0,7]$, $[9,0,5,3]$, $[7,9,4,3]$, $[0,7,9,0,5]$, $[3,0,7]$, and $[9,4,3]$ [8]. To understand what these meant, we made the robot move by using these keywords. We found that $[7,9,4,3]$ and $[9,4,3]$ corresponded to the up-down motion. The difference between these two motions was that $[7,9,4,3]$ made the robot place its hand on the table before it raised its hand. In contrast, $[9,4,3]$ made the robot directly raise its hand. The motion generated by $[7,9,4,3]$ is shown in **Fig. 11**. Moreover, we found that $[7,9,0,5,3,0,7]$ corresponded to the square motion, and $[9,0,5,3,0,7]$ and $[9,0,5,3]$ were parts of the square motion. This meant that the keyword extractor could not completely eliminate substrings of keywords. The motion generated by $[7,9,0,5,3,0,7]$ is shown in **Fig. 12**. $[3,0,7]$ was the motion for remaining in the base position and moving slightly to the right. The experimenter unconsciously seemed to keep the pole at the lower right and then move right before he raised his hand. The movement where the experimenter moved his hand to the lower right and then moved it to the right was extracted as a key motion.

However, it was difficult for the experimenter to translate the motion generated by $[0,7,9,0,5]$. These sorts of unintended results were possible to acquire in our framework because our model was a self-organizing learning architecture, and did not utilize any direct teacher signal in the segmentation. The motion generated by $[0,7,9,0,5]$ is shown in **Fig. 13**.

After acquisition, we made the robot demonstrate these motions five times. When the robot exhibited meaningful motions, i.e., $[7,9,0,5,3,0,7]$ (square motion), $[7,9,4,3]$, or $[7,4,3]$ (up-down motion), we had the experimenter demonstrate up-down motion as a response to the robot's motion. However, the experimenter ignored the robot's other motions by not responding in any way.

After the robot had demonstrated motion, we made it observe the experimenter's motion for about five seconds. Note that even when the experimenter did not return an up-down motion to the robot's keywords, a time series containing the user's meaningless motion was displayed to the robot.

8. Each numeral represents a hidden state of SARM.

If an up-down motion was returned to the robot in reply to its motion, the robot could usually recognize the motion as a keyword, $[7,9,4,3]$ or $[7,4,3]$, and it got to know that its exhibited motion was meaningful to the experimenter because he gave some meaningful feedback to the robot. In contrast, when the robot demonstrated other motions to the experimenter, it did not receive any meaningful responses. By using these experimenter's reactions, the robot could filter out meaningless motions from the acquired keywords, and it finally acquired two meaningful motions, i.e., an up-down motion and a square motion, which the experimenter considered to be meaningful throughout these human-robot interactions.

These results indicate that our imitation-learning framework enabled the robot to acquire several motions, recognize them, and identify various keywords that were meaningful to the experimenter. In this framework, we assumed that the meaningless motions demonstrated by the robot did not cause the experimenter to exhibit any meaningful motions.

The robot was also able to learn that it should display the up-down motion represented by $[7,9,4,3]$ or $[7,4,3]$ after the experimenter had demonstrated an up-down motion or a squared motion. This is the simplest process in role-reversal imitation. Therefore, our learning framework achieved the simplest computational role-reversal imitation through unsegmented human-robot interactions.

In this experiment, we evaluated the efficacy of our model by watching the motions the robot had acquired. Although it would have been important to evaluate how well the learning architecture performed quantitatively, not qualitatively, this type of algorithm to extract motion primitives does not currently have appropriate quantitative-evaluation criteria. The learning task, after the times series are segmented and target segmented time series are collected, represents conventional supervised learning. Therefore, we evaluated the performance of the algorithm qualitatively. In this paper, we employed SARM whose learning algorithm is derived based on the EM algorithm to ensure that the algorithm at least obtained a suboptimal learning result. We did not discuss SARM itself in this paper. If the target mathematical model for the extraction task had been defined, methods of extraction could have been fairly compared. However, the true segmentation of motion primitives, or meaningful key motion of human continuous motion, is usually unknown. The existence of such true segments is also doubtful. The performance of most application examples of extraction algorithms seems to have been evaluated qualitatively thus far. Our future work is to construct appropriate evaluation criteria for role-reversal imitation.

4. Conclusion

We described a computational model of role-reversal imitation learning through unsegmented human-robot interactions. Our experiment revealed that a robot embedded with our learning architecture could acquire a few motions from the continuous motion data provided by a human experimenter. In most previously proposed methods of imitation learning, users interacting with autonomously learning robots explicitly taught them some motions. Their target time series of imitation learning had to be finely segmented. However, our model made a robot learn several motions through unsegmented continuous human-robot interactions. This means our proposal will not only create a great leap forward in the research field of computational imitation learning, but should also produce novel experiences in human-robot interactions because our model enable us to create an autonomous robot, e.g., a robot for entertainment that can acquire several motions through continuing natural human-robot interactions. This should enable owners of entertainment robots to literally nurture them. In addition to this, we proved that the simplest role-reversal imitation through continuous human-robot interactions could be accomplished computationally. This represents qualitative progress in the field of robotic imitative learning.

However, our model still involves too many heuristics and manually determined parameters. For example, we manually determined the number of hidden states, and four parameters in the keyword extractor. In future work, we intend to apply a method of model selection to determine an adequate number of hidden states and study how to determine the parameters in the keyword extractor. At present, the model of the keyword extractor does not have a theoretical basis in information theory in contrast to SARM, which has been theoretically formulated. Therefore, model selection criteria for role-reversal imitation are required. In addition to this, three learning processes, i.e., SARM, keyword extraction, and user reaction filtering are mathematically separated. Additionally, we have to understand what the total learning architecture should optimize through a total learning process. This is also a problem. To overcome these problems, we intend to formulate this learning architecture as a mathematically sophisticated total-learning architecture in future work. We should be able to derive adequate evaluation criteria and model-selection and parameter-tuning methods by integrating separate learning processes.

One of the most important assumptions in our model is that the distribution of key phrases in documents generated by segmenting natural human motion with SARM has similar features to the distribution of keywords in written documents. We have assumed that natural motion by humans satisfies this assumption. However, if this assumption is not satisfied, our framework will not work. To ensure our framework will work, it is necessary to study the structure of human motion generated in human-robot or human-human interactions from the viewpoint of behavioral science. In addition, we tested our framework in a three-dimensional world, which we discussed in this paper. However, human bodies have numerous degrees of freedom. To apply this framework to a humanoid robot, we have to examine the scalability of this model.

Some might point out that our framework constrains the way user's react with the robot. We were also concerned that the learning process may have affected natural human-robot interactions. However, our method does not require the human user to stop or push a button in the first and second phase, i.e., "learning SARM" and "keyword extraction." Although we acknowledge that our method could not achieve completely natural human-robot interactions through imitation learning, we believe it could contribute to natural human-robot interactions. Also, when a human child understands a new motion, he or she has to utilize some information fed back to the child to statistically characterize his or her motion. From the viewpoint of a constructive approach to human imitation learning, his or her mother has to constrain her reaction to leading the child's learning process. Therefore, we think constraining a user's reactions does not directly result in unnatural interactions. Based on this concept, we assumed that a user would exhibit meaningful motion after a robot had exhibited meaningful motion. This assumption is based on role-reversal imitation (see **Fig. 1**). However, human users possibly use other teaching strategies. For example, when a robot exhibits a meaningless motion, he or she might re-exhibit a similar meaningful motion. In such cases, our model does not work well. In this paper, we focused on "rolereversal imitation." Therefore, we assumed a single interaction to achieve the learning process. Further studies are required to treat various learning processes.

Finally, our model should enable autonomous robots to learn a variety of behaviors through continuous interactions with human users. This should enrich human-robot interactions qualitatively. Our method will potentially have a huge impact not only on the field of robotic imitation learning, but also on the field of human-robot interactions. Unfortunately, this paper has left several questions unanswered as was discussed earlier. However, our approach has sufficient qualitative novelty and is progressive in comparison with conventional robotic imitation learning.

Acknowledgements

This paper was partially supported by the Center of Excellence for Research and Education on Complex Functional Mechanical Systems (The 21st Century COE program of the Ministry of Education, Culture, Sports, Science and Technology, Japan), by a research grant from the National Institute of Informatics, and by the Ministry of Education, Culture, Sports,Science and Technology, Japan, Grant-in-Aid for JSPS Fellows, 17-1685 and 19-3467.

References:

[1] M. Asada, K. MacDorman, H. Ishiguro, and Y. Kuniyoshi, "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," Robotics and Autonomous Systems, Vol.37, Nos.2-3, pp. 185-193, 2001.

[2] A. Billard, Y. Epars, S. Calinon, S. Schaal, and G. Cheng, "Discovering optimal imitation strategies," Robotics and Autonomous Systems, Vol.47, Nos.2-3, pp. 69-77, 2004.

[3] B. Chiu, E. Keogh, and S. Lonardi, "Probabilistic discovery of time series motifs," Proc. of the ninth ACM SIGKDD Int. Conf. on Knowledge discovery and data mining, pp. 493-498, 2003.

[4] K. Church, "Empirical estimates of adaptation: the chance of two noriegas is closer to p/2 than p 2," Proc. of the 17th Conf. on Computational linguistics, Vol.1, pp. 180-186, 2000.

[5] P. Dunham and F. Dunham, "Effects of mother-infant social interactions on infants' subsequent contingency task performance," Child Development, Vol.61, No.3, pp. 785-793, 1990.

[6] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," Int. Journal of Robotics Research, Vol.23, No.4, pp. 363-377, 2004.

[7] M. Ito, K. Noda, Y. Hoshino, and J. Tani, "Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model," Neural Networks, Vol.19, No.3, pp. 323-337, 2006.

[8] J. Lin, E. Keogh, S. Lonardi, and P. Patel, "Finding Motifs in Time Series," Proc. of the 2nd Workshop on Temporal Data Mining, pp. 53-68, 2002.

[9] C. Moore, "Joint Attention: Its Origins and Role in Development," Lawrence Erlbaum Assoc Inc., March, 1995.

[10] K. Murphy, "Switching Kalman filters," Dept. of Computer Science, University of California, Berkeley, Tech. Rep., 1998.

[11] C. Nehaniv and K. Dautenhahn, "The correspondence problem," Imitation in Animals and Artifacts, pp. 41-61, 2002.

[12] M. Okada, K. Tatani, and Y. Nakamura, "Polynomial design of the nonlinear dynamics for the brain-like information processing of whole body motion," Robotics and Automation, 2002, Proc., ICRA'02, IEEE Int. Conf. on, Vol.2, 2002.

[13] K. Sugiura and N. Iwahashi, "Learning Object-Manipulation Verbs for Human-Robot Communication," Workshop on Multimodal Interfaces in Semantic Interaction, Int. Conf. on Multimodal Interfaces, 2007.

[14] Y. Tanaka, K. Iwamoto, and K. Uehara, "Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle," Machine Learning, Vol.58, No.2, pp. 269-300, 2005.

[15] T. Taniguchi and T. Sawaragi, "Incremental acquisition of behaviors and signs based on a reinforcement learning schemata model and a spike timing-dependent plasticity network," Advanced Robotics, Vol.21, No.10, pp. 1177-1199, 2007.

[16] M. Tomasello, "The Cultural Origins of Human Cognition," Harvard Univ Pr, reprint edition, March, 2001.

[17] K. Umemura, "Related Word-pairs Extraction without Dictionaries," Technical report, IPA Exploratory Software Project development result report, 2000 (in Japanese) http://www.ipa.go.jp/archive/NBP/12nendo/12mito/mdata/10-36h/10-36h.pdf

[18] J. Watson, "Contingency perception in early social development," Social perception in infants, pp. 157-176, 1985.

[19] R. Yokoya, T. Ogata, J. Tani, K. Komatani, and H. Okuno, "Experience Based Imitation Using RNNPB," Intelligent Robots and Systems, 2006 IEEE/RSJ Int. Conf. on, pp. 3669-3674, 2006.

Name:
Tadahiro Taniguchi

Affiliation:
Assistant Professor, Department of Human & Computer Intelligence, Ritsumeikan University

Address:
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577, Japan

Brief Biographical History:
2006- JSPS Postdoctoral fellow, Graduate school of Engineering, Kyoto University
2007- JSPS Postdoctoral fellow, Graduate school of Informatics, Kyoto University
2008- Assistant Professor, Department of Human & Computer Intelligence, Ritsumeikan University

Main Works:
• T. Taniguchi and T. Sawaragi, "Incremental Acquisition of Behaviors and Signs based on Reinforcement Learning Schema Model and STDP," Advanced Robotics, Vol.21, No.10, pp. 1177-1199, 2007.
• T. Taniguchi and T. Sawaragi, "Incremental acquisition of multiple nonlinear forward models based on differentiation process of schema model," Neural Networks, Vol.21, No.1, pp. 13-27, 2008.

Membership in Academic Societies:
• The society of instrument and control engineers (SICE)
• Japanese Neural Network Society (JNNS)
• The Japanese Society for Artificial Intelligence (JSAI)

Name:
Naoto Iwahashi

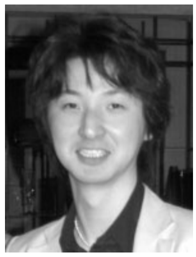
Affiliation:
National Institute of Information and Communication Technology
Advanced Telecommunications Research Institute Int.

Address:
2-2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-0288, Japan

Brief Biographical History:
1985-1998 Sony Corporation
1998-2003 Sony Computer Science Laboratories
2003- Advanced Telecommunications Research Laboratories Int.
2006- National Institute of Information and Communications Technology

Main Works:
• "Robots that Learn Language: A Developmental Approach to Situated Human-Robot Conversation," N. Sanker (Ed.) Human-Robot Interaction, I-Tech Education and Publishing, pp. 95-118, 2007.

Membership in Academic Societies:
• Japanese Society for Artificial Intelligence (JSAI)
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• Japanese Cognitive Science Society (JCSS)



Name:
Komei Sugiura

Affiliation:
Researcher, ATR Spoken Language Communication Research Laboratories
Expert Researcher, Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology

Address:
2-2-2 Hikaridai, Seikacho, Sorakugun, Kyoto 619-0288, Japan

Brief Biographical History:
2007- JSPS Postdoctoral fellow, Graduate school of Informatics, Kyoto University
2008- Researcher, ATR Spoken Language Communication Research Laboratories

Main Works:
• "Co-evolution of Sensors and Controllers," IEEJ Trans. EIS, Vol.124-C, pp. 1938-1943, 2004.

Membership in Academic Societies:
• The society of instrument and control engineers (SICE)
• The Robotics Society of Japan (RSJ)
• The Japanese Society for Artificial Intelligence (JSAI)

Name:
Tetsuo Sawaragi

Affiliation:
Professor, Dept. of Mechanical Engineering and Science, Graduate School of Engineering, Kyoto University

Address:
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan

Brief Biographical History:
1981 Received B.S. degree in Systems Engineering from Kyoto University
1983 Received M.S. degree in Systems Engineering from Kyoto University
1986-1994 Research Associate, Department of Precision Mechanics, Faculty of Engineering, Kyoto University
1988 Received Ph.D. degree in Systems Engineering from Kyoto University
1991-1992 Visiting Scholar, Dept. of Engineering-Economic Systems, Stanford University, USA
1994-2002 Associate Professor, Department of Precision Mechanics, Faculty of Engineering, Kyoto University
2002-2005 Professor, Department of Precision Mechanics, Graduate School of Engineering, Kyoto University
2005- Professor, Department of Mechanical Engineering and Science, Graduate School of Engineering, Kyoto University

Main Works:
• "Self-Reflective Segmentation of Human Bodily Motions Using Recurrent Neural Networks," IEEE Trans. on Industrial Electronics, Vol.50, No.5, pp. 903-911, 2003.
• "Human-Machine Collaborative Knowledge Creation: Capturing Tacit Knowledge by Observing Expert's Demonstration of Load Allocation," Int. Journal of Knowledge and Systems Sciences, Vol.3, No.2, pp. 9-19, 2006.

Membership in Academic Societies:
• The Institute of Systems, Control and Information Engineers (ISCIE)
• Human Interface Society (HIS)
• Japan Society for Fuzzy Theory and Systems (SOFT)
• Japanese Society for Artificial Intelligence (JSAI)
• The Japan Society of Mechanical Engineers (JSME)
• The Institute of Electrical and Electronics Engineers (IEEE)