Research Subject

Stochastic Information Processing that unifies Recognition and Generation of Motion Patterns –Toward Symbolical Understanding of the Continuous World– (Nakamura Group)

(1) Goal and summary

The purpose of the research is to propose a novel brain-like information processing framework which can connect motion patterns and symbols. We have focused on two knowledge: "Mimesis Theory" and "Mirror Neurons" for the purpose. The discovery of mirror neurons[1] have been a notable topic of brain science which have been found in primates' brain and humans' brain, fire when the subject observes a specific behavior and also fire when the subject start to act the same behavior. Furthermore, it is located on Broka's area which has close relationship between language management. The fact suggests that the behavior recognition process and behavior generation process are combined as the same information processing scheme, and the scheme is nothing but a core engine of symbol manipulation ability. Indeed, in Donald's "Mimesis Theory"[2], it is said that symbol manipulation and communication ability are founded on the behavior imitation, that is integration of behavior recognition and generation. We believe that a paradigm can be proposed taking advantage of the mirror neurons, with considerations of Deacon's contention[3] that the language and brain had evolved each other.

For the purpose, we have developed following four frameworks: (1) Mutual connection model between motion patterns and symbols based on hidden Markov model, (2) Keyframe compression and decompression for time-series data based on the continuous hidden Markov model, (3) Imitation learning model with embodiment based on discrete continuous hybrid HMM, (4) Development and manipulation of proto-symbols based on geometric proto-symbol space.

Mutual connection model between motion patterns and symbols based on hidden Markov model

We have focused on a stochastic information processing framework of the hidden Markov model(HMM) in order to integrate symbol representation and motion patterns of humanoid robots which have a lot of degree of freedoms. The HMM is regarded as a symbol representation which is named as "proto-symbol", also used for the development of the mutual connection model . The mutual connection model consists of two phases. In the first half, observed motions are transformed into motion elements by comparing, and are abstracted as proto-symbols. Observed motion patterns are analyzed into the motion elements, and the sequence of the motion elements are abstracted into proto-symbols, regarded as a series of behavior. Figure 1 shows the HMM which expresses motion patterns, that is time-series data of motion elements.

Generation of motion patterns from the HMM is equal to generate time-series data of motion elements. However, it is difficult to calculate the sequence by only the HMM, because the generation process is equal to search a motion pattern which has the best likelihood value among the all the entire motion patterns. Most simple way to generate suitable motion patterns is to find the maximum likelihood by scanning the entire pat-

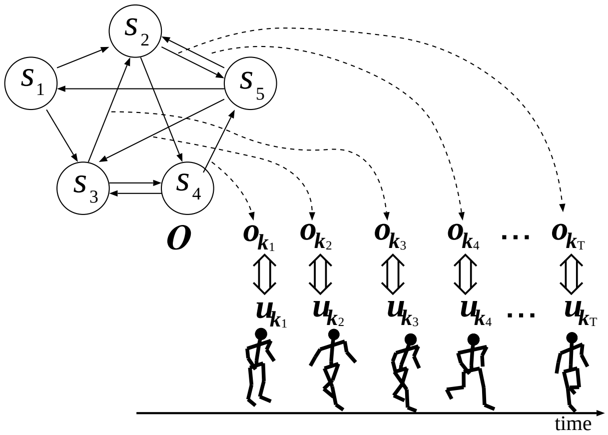

Figure 1: Representation of motion patterns using hidden Markov model

tern space. However, it is difficult to adopt this method because the size of the search space will be increase in proportion to the exponential of the time length of the motion pattern. In order to encode a motion pattern into a chromosome, each motion element is corresponded to each gene. As the fitness of the chromosome, the likelihood of that the motion patterns are generated by the HMM is used. It have also adopted translocation not simple crossover and mutation. It is suitable for the evolution to keep a series of behavior because the block of self motion elements indicates the series of behavior.

A mathematical model for the integration of motion recognition and generation is achieved using above methods. Figure 2 shows an outline of the proposed framework.

Keyframe compression and decompression for time-series data based on the continuous hidden Markov model

Memory of motion patterns as data, comparison of a new motion pattern with the data, and playback of one from the data are inevitably involved in the information processing of intelligent robot systems. Such computation forms the computational foundation of learning, acquisition, recognition, and generation process of intelligent robotic systems. Motion patterns along with temporal sensory data would be appropriate to describe behaviors of a robot. This is the computational problem of time series data and the subject of the present paper.

The computational problem of time series data would need to consider: (1) efficiency of data compression/decompression, and (2) unification of algorithms for memory (compression), comparison and playback (decompression). The former is mandatory since it determines the volume of database of motion patterns. The latter is not a must, but an important requirement to maintain consistency of the three kinds of computation. We have focused on the keyframe representation for the purpose. Keyframe is one of the famous method for motion design, especially in computer graphics, which is a motion patterns at several impressive moments. Time-series data of motion patterns are combined using these keyframe in the computer graphics. Recently, The keyframe is used for the motion planning of robots and motion recognition because the frame representation has affinity for such issues.

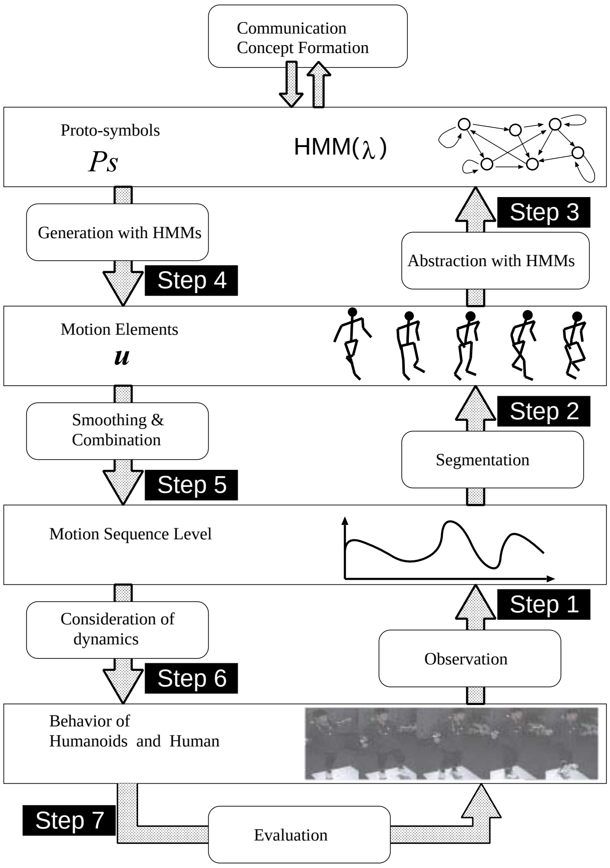

Figure 2: Mutual connection model between motion patterns and symbols based on hidden Markov model

Proposed mimesis model has no criterion for designing the motion element. It is effective for the mimesis model to adopt the keyframe representation for the motion elements design.

For the keyframe representation, continuous hidden Markov model (CHMM) is used as shown in Fig.3.The CHMM consists of a finite set of states $\boldsymbol{S} = \{s_1, \cdots, s_N\}$, a state transition probability matrix $\boldsymbol{A} = \{a_{ij}\}$, output probability functions $\boldsymbol{b} = \{b_i(\boldsymbol{o})\}$ and an initial distribution vector $\boldsymbol{\pi} = \{\pi_i\}$, that is a set of parameter $\boldsymbol{\lambda} = \{\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{b}, \boldsymbol{\pi}\}$. $b_i(\boldsymbol{o})$. Here, b_i is output probability density function

$$b_i(\boldsymbol{o}) = \sum_{j=1}^{M} c_{ij} \mathcal{N}_{ij}(\boldsymbol{o}; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_{ij}) \tag{1}$$

, that relates continuous output vector $\boldsymbol{o}$ with the i-th state node s_i. M indicates the number of Gaussian functions as $\mathcal{N}(\boldsymbol{o}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The CHMM generates motion patterns by the stochastic process as shown in Fig.3.

We have defined the motion elements as keyframes of the motion pattern as follows:

$$\boldsymbol{u} \stackrel{\text{def}}{=} \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}. \tag{2}$$

These parameters is calculated by Baum-Welch algorithm.

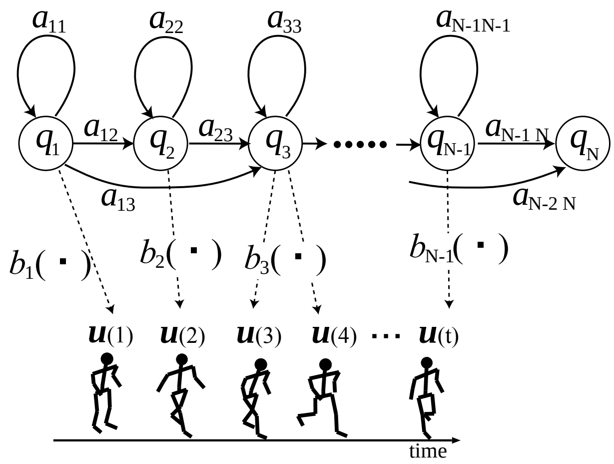

Figure 3: Motion pattern representation using continuous HMM

Here, there are sway among each reproduction processes because of the property of stochastic models. The time length of the motions T_i always changes by the state transition probability $\boldsymbol{A}$, the value for each moment of the time-series data also always changes by the output probability $b_i(\boldsymbol{o})$. We propose an average strategy in order to cancel these sway.

$\boldsymbol{step1}$ Getting a state transition sequence $\boldsymbol{Q} = \left[s_{k[1]}, s_{k[2]}, \cdots, s_{k[T]}\right], (k[i] \in \{1, 2, \cdots, N\})$ with a trial of the state transition process.

$\boldsymbol{step2}$ An average of transition sequence $\hat{\boldsymbol{Q}}$ is calculated by n_q times repetition $(\boldsymbol{Q}_1, \cdots, \boldsymbol{Q}_{n_q})$ of the $\boldsymbol{step1}$.

$\boldsymbol{step3}$ Output time-series data $\boldsymbol{O}$ is calculated by a trial of output according to the average transition sequence $\hat{\boldsymbol{Q}}$.

$\boldsymbol{step4}$ $\boldsymbol{O}_1, \cdots, \boldsymbol{O}_{n_o}$ is calculated by n_o times repetition from $\boldsymbol{step1}$ to $\boldsymbol{step3}$.

$\boldsymbol{step5}$ Average time-series data $\hat{\boldsymbol{O}}$ is calculated by the $\boldsymbol{O}_1, \cdots, \boldsymbol{O}_{n_o}$ after regularization of the time length.

where n_q, n_o are decided experimentally.

Figure 4 shows the example decomposition result using the method. Target data is a certain joint angle of a humanoid robot. The dot-line indicates the original time-series data, the dashed line indicates a result of single output trial ($\boldsymbol{Q}_i$). There are deviance for time direction and value direction, however, the deviance for time direction was cancelled using $\boldsymbol{Step2}$, and the deviance for value direction was also cancelled using $\boldsymbol{step5}$. Final output result is shown as solid line. As the figure shows, output time-series data is similar to the original time-series data.

Imitation learning model with embodiment based on discrete/continuous hybrid HMM

There were several remained problems in the mimesis model as follows:

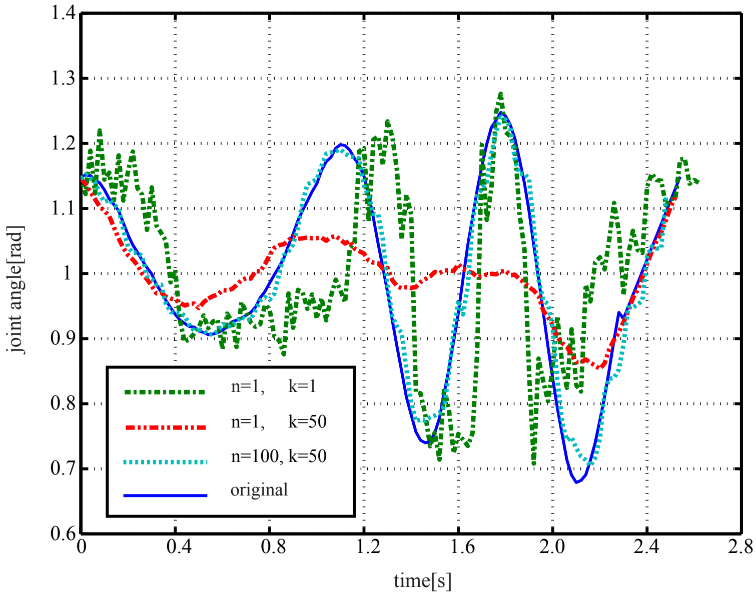

Figure 4: Generation of time-series data using continuous HMM

1. There is no exact principle how to design the motion elements. In previous works, static and limited motion elements had been embedded by the developer beforehand. It is necessary for the mimesis system to develop the motion elements in order to be suitable for imitation learning from observation experience.

2. Physical condition of the motion had not been taken into consideration. As physical characteristic of learner and demonstrator is different, therefore, the observer cannot reproduce the same motion. The motion elements have to be suitable for both recognition of other's motion and embodiment of humanoids.

3. The motion elements were correspond to each joint. Thus the number of motion elements becomes no less than the number of DOFs in order to represent the whole body motion. it causes the complexity of symbol representation. Furthermore, the motion elements should represent correlation information between each joint.

Because of the above-mentioned reason, we propose a new method for acquisition of motion elements with two characteristics; a) use of continuous HMM and b) modification of elements during observation and generation loop. The approach a) enables the system to avoid the problems 1) and 2). On the other hand the approach b) enables the system to avoid the problem 2) and 3).

For the purpose, we have introduced continuous HMM which is a kind of HMM which can treat continuous data. The difference between DHMM and CHMM is that the transition process outputs continuous vectors as shown in Fig.3.

Although many advantages are available, CHMM have a disadvantage that huge computational quantity is needed. It should take much time for motion generation and recognition. Therefore we have proposed a hybrid Hidden Markov Model which consists of CHMM and DHMM as shown in Fig.5. In motion recognition and generation phase, DHMM are used which computational quantity is little. In motion elements acquisition phase, CHMM are used which computational quantity is large.

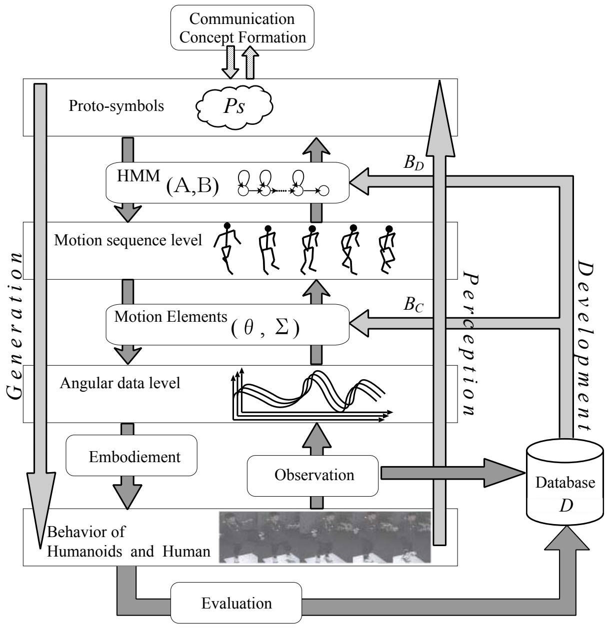

Figure 5: Mimesis model based on discrete continuous hybrid HMM

The mimesis model acquires adequate motion elements during the repetition of motion recognition and generation. Following steps are the procedure for the motion elements development.

1. Generating of the motion patterns from proto-symbols.
2. Evaluation of the motion patterns based on recognition criterion.
3. If the score is good, the motion patterns is added into the database.
4. Returning to the step 1, after the recalculation of motion elements.

The performance of motion recognition and generation is influenced by the characteristic of motion elements. If the motion elements had no relationship between the observed motion, the recognition process would be failed. Therefore, we adopted an approach that the system searches the best motion elements with an evaluation criterion whether the generated motion would be fit for the body and the recognition would be succeeded against familiar motion. Using the method, the humanoid can acquire adequate motion elements through repetition of motion perception and generation.

A result with the limitation condition is shown in Fig.6. In the figure, three axes indicate hip joint (pitch), knee joint and ankle joint (pitch). The curved line in the figure corresponds to the motion trajectory. The dots indicate acquired motion elements. The result shows that the motion elements are basically located near by the original motion trajectory. Additionally, the Motion elements are gathered not only on the A area, but also on the B area in Fig.6. These motion elements located on the B area is acquired by the generated self-motions in the database, which fits for the humanoid embodiment. This result shows that the both motion elements are acquired; elements for the recognition of others' motions (A area) and ones for the generation of self-motion (B area).

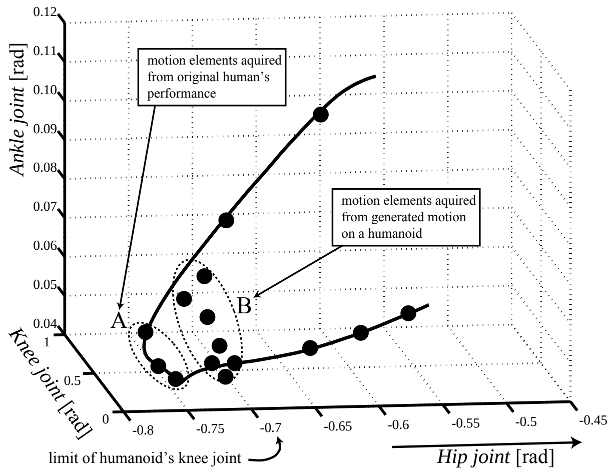

Figure 6: Acquired motion elements

Development and manipulation of proto-symbols based on geometric proto-symbol space

Symbols are required to represent similarity or opposite between each symbol, however, proto-symbol representation cannot express the relationship between each proto-symbol as shown in Fig.7(left). Therefore, we have extended our HMM based method in order to express a geometric proto-symbol space which contains relative distance information among proto-symbols[6] as shown in Fig.7(right).

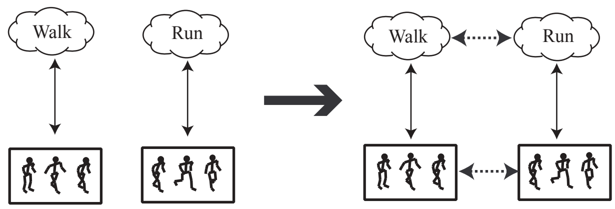

Figure 7: Usual mimesis model (left) and advanced mimesis model (right)

Through distance is needed for construction of space, distance between two HMM is not able to be defined easily because it is stochastic model. For such stochastic modes, there is a method in order to express the distance information. In this paper, we adopt Kullback-Leibler information as the representation of distance between HMM. To say strictly, the Kullback-Leibler information is not distance because it does not satisfy the property of the distance; triangle inequality and symmetry, therefore we call the Kullback-Leibler information as degree of similarity of the HMM. The Kullback-Leibler information against two HMM λ_1 and λ_2 is defined as follows [4].

$$D(\lambda_1, \lambda_2) = \frac{1}{n} \sum_i \frac{1}{T_i} \left[\log p(\boldsymbol{y}_1^{T_i} | \lambda_1) - \log p(\boldsymbol{y}_1^{T_i} | \lambda_2) \right] \tag{3}$$

$$D_s(\lambda_1, \lambda_2) = \frac{1}{2} \left(D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1) \right) \tag{4}$$

where, $\boldsymbol{y}_1^{T_i}$ is time-series data for learning of the λ_1 whose time length is T_i, n is the number of observed time-series data.

In order to construct proto-symbol space from the distance information, multidimensional scaling (MDS) is used. MDS is a method that accepts distance information among elements and outputs position of each element in the generated space. Let the similarity between i-th element and j-th element as f_{ij}, the distance between i-th and j-th element as d_{ij}. MDS makes the following error to be minimum for the space construction. We have adopted $D_s(\lambda_i, \lambda_j)$ as the f_{ij}.

motion recognition and generation in the proto-symbol space In case that an unknown behavior is observed by the mimesis model which have existing 6 proto-symbols in the geometric space, observed motion pattern is converted into a parameter of HMM $\hat{\lambda}$. Next, the parameter $\hat{\lambda}$ is projected in the proto-symbol based on the distance calculation result between $\hat{\lambda}$ and each existing proto-symbols $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$.

Using the geometric proto-symbol space, the model can recognize unknown behaviors as a state point in the proto-symbol space, and generate novel behaviors using combination of proto-symbols by known proto-symbols, that is geometric proto-symbol manipulation in the proto-symbol space.

Creating a novel proto-symbol is equal to create a novel state point on the proto-symbol space. To create a novel state point, following composition regulation is used:

$$\begin{aligned} b_i(\boldsymbol{o}) = & \sum_{m=1}^{M} \alpha c_{im_A} N(\boldsymbol{\mu}_{im_A}, \boldsymbol{\sigma}^2_{im_A}) \\ & + \sum_{m=1}^{M} (1-\alpha) c_{im_B} N(\boldsymbol{\mu}_{im_B}, \boldsymbol{\sigma}^2_{im_B}) \end{aligned} \tag{5}$$

$$a_{ij} = \alpha a_{ij_A} + (1-\alpha) a_{ij_B} \tag{6}$$

Eq.(5) and Eq.(6) is applied when a novel state point is located on a straight line connecting the two points (λ_A and λ_B). When a novel state point is not on any straight lines connecting known proto-symbols, the parameter is composed according to the distance ratio among each known proto-symbol as follows:

$$\begin{aligned} b_i(\boldsymbol{y}) &= \sum_{m=1}^{M} \frac{1}{d_l \sum_l \frac{1}{d_l}} c N(\boldsymbol{\mu}^l_{im}, {\boldsymbol{\rho}^l_{im}}^2) \\ a_{ij} &= \sum_{m=1}^{M} \frac{1}{d_l \sum_l \frac{1}{d_l}} a^l_{ij} \end{aligned} \tag{7}$$

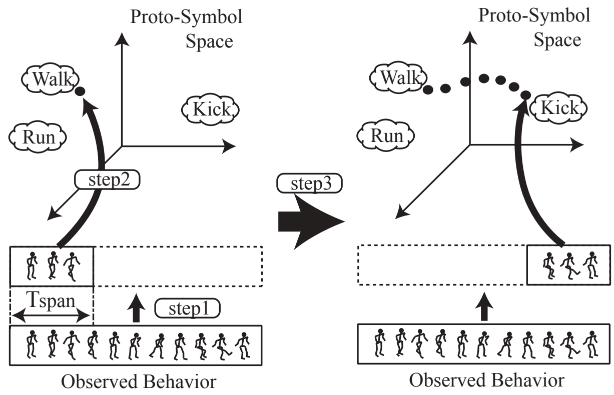

Figure 8: Procedure of projecting motion in proto-symbol space

where d_l is the distance between a novel state point and known proto-symbol λ_l. Finally, law-level motion pattern is generated from the state point.

Humans' behavior are always novel and unknown motion, therefore, motion recognition system always have to output an unknown symbol representation which corresponded to the novel behavior. The proto-symbol space can indicate a long motion pattern as a sequence of state points in the geometric space, and also generate a long motion pattern from the sequence of state points in the proto-symbol space.

The outline of novel motion recognition as state sequence in the proto-symbol space is shown if Fig.8 In step 1, focusing on the period of time T_{span} in the observed motion pattern $\boldsymbol{O} = \left[\boldsymbol{o}_1 \ \boldsymbol{o}_2 \ \cdots \ \boldsymbol{o}_T\right]$. Let the cut off motion pattern be $\boldsymbol{O}_1 = \left[\ \boldsymbol{o}[1] \ \boldsymbol{o}[2] \ \cdots \ \boldsymbol{o}[T_{span}]\ \right]$. In step 2, state point is decided using mentioned method in previous paragraph. Next, shift the focus point, and let the k-th focus point be $\boldsymbol{O}_k = \{\boldsymbol{o}_{1+(k-1)\cdot T_{step}}, \cdots, \boldsymbol{o}_{1+T_{span}+(k-1)\cdot T_{step}}\}$, with increase the index as $k = 1, 2, \cdots, \frac{T-1-T_{span}}{T_{step}} + 1$. Finally in step 3, sequence of state point in the proto-symbol space is acquired.

Furthermore, motion generation can be performed against the state transition sequence in the proto-symbol space. In this paragraph, a motion generation method in which a state transition sequence have been given. Let be the state transition as $\boldsymbol{x}[1], \boldsymbol{x}[2], \cdots, \boldsymbol{x}[n]$. As a generation method in which a fixed state point is given in the proto-symbol space is introduced before, the continuous generation by transitional state points is equivalent to the average of motions which generated by those state points.

Figure 9 shows the outline of the generation process. In step 1, motion patters are generated from each state point in the proto-symbol space using the proposed method. In step 2, the time length of all motion patterns are set to the same value T_c in order to composition. In step 3 and 4, partial motion patterns are picked up based on the phase information for each state point, that is, charging period of time for each state point. Finally in step 5, composite motion pattern are generated.

Result of motion generation and recognition using the hierarchical mimesis model
We confirmed the performance of the proposed space construction method against six kinds of motion; walk, kick, squat, stoop, stretch and throw, as shown in Fig.10. At first, we gave 10 dimensional vector for each $\{x_1, x_2, \ldots, x_n\}$. As from first to fourth dimen-

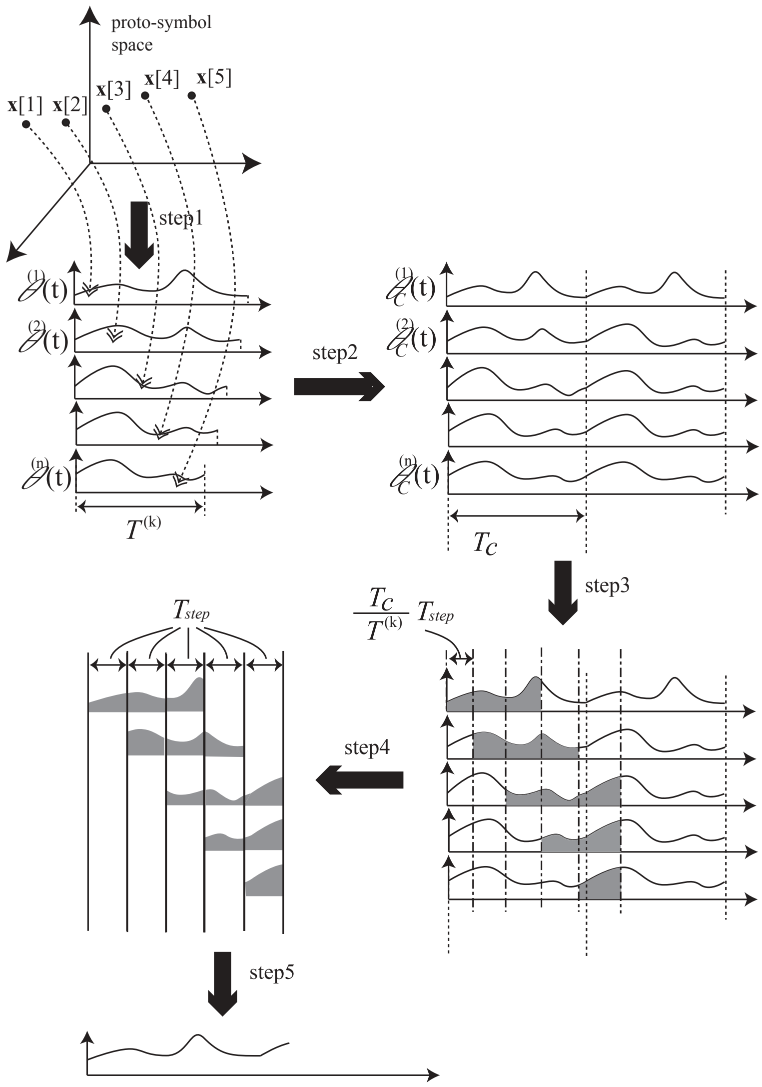

Figure 9: Procedure of motion pattern generation in proto-symbol space

sions are effectively used for the space construction, however, the rest of the dimension are not well used. Therefore, we adopted three dimensional proto-symbol space as shown in Fig.12.

We performed a recognition experiment in which novel motion pattern are transfered into a sequence of state points in the proto-symbol space. The target motion is “walking first, then shift to kicking”. The result is shown in Fig.13. As the diagram shows, recognized dot marks starts from the proto-symbol of “walk”, ends at the proto-symbol of “kick”.

Motion recognition using the HMM is famous method, thus many research are proposed for gesture recognition or behavior understanding [5], however, no research has been existed in which motion is generated from HMM.

In this experiment, we have investigate the motion output when a trajectory $\boldsymbol{X}[t]$ is given in the proto-symbol space. As a given trajectory, we prepared simple line trajectory from the “walking” state point to the “kicking” state point. Figure 13 is the result of

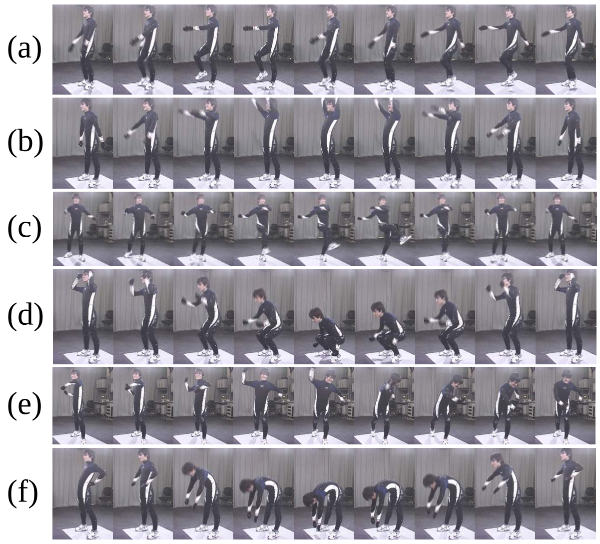

Figure 10: Six motions performed by human

Figure 11: A novel motion pattern : “kick after walk”

motion output. As the figure shows, motion of the humanoid is adequately controlled as the symbol manipulation in the proto-symbol space.

Approach to intelligence based on hierarchical structure In this model, motion patterns of humans and humanoids $\boldsymbol{\theta}(t)$ are abstracted by HMM, and converted as a state point $\boldsymbol{x}$ in the proto-symbol space. A static state point $\boldsymbol{x}$ is corresponded to a time-series data $\boldsymbol{\theta}(t)$. As the state point is a vector in the proto-symbols space, time-series data of the state point $\boldsymbol{X}[t] = \left[\ \boldsymbol{x}[1]\ \boldsymbol{x}[2]\ \cdots\ \boldsymbol{x}[n]\ \right]$ also can be abstracted using HMM. We call such a tautological HMM as “hierarchical mimesis model”. We also define the abstracted representation from the state point sequence $\boldsymbol{X}[t]$ as meta proto-symbol as shown in Fig.14

Using the hierarchical mimesis model, motion patterns $\boldsymbol{\theta}(t)$ are converted to more abstract representation $\mathcal{M}$ by the repetition of abstract process. One of the advantage of adopting the HMM is easiness for the connection between low-level motion patterns and

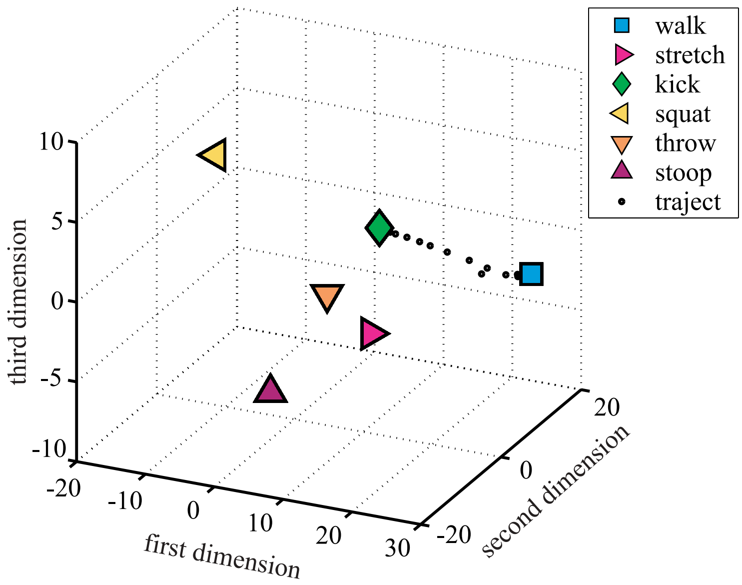

Figure 12: A result of proto-symbol construction and recognition result using it

high-level symbols.

(2) Results and their importance

In this research, we have proposed a framework which realizes motion recognition/generation, symbolize the motion patterns, and leads to intelligence of humanoid based on mimesis theory. In our mimesis model, proto-symbols and motion elements are introduced with hidden Markov model in order to integrate following four abilities using only one mathematical model; (1) abstraction of motion patterns and symbol representation , (2) acquisition of motion elements, that is keyframe representation of motion patterns based on continuous HMM, (3) imitation learning based on discrete/continuous hybrid HMM, and (4)recognition of unknown motion and generation of novel motion based on a geometric proto-symbol space. Through experiences, the feasibility of proposed mimesis model is cleared. Furthermore, we proposed hierarchical mimesis model in which the HMM are overlapped in order to connect low level motion patterns and high level symbols.

Symbol emergence had been tried in conventional research of artificial intelligence. The most difficult issue of the symbol emergence is how to manipulate the created symbol representation, contrary to the easiness of symbol creation. Deacon have proposed the symbol development model [3] as shown in Fig.15. In his theory, symbolic representation is developed from indexical level and iconic level. In the indexical level, simple relationship between a motion pattern and a symbol representation is established, however, relationship between each symbol and motion pattern is not considered. In the transitional level, relationship between each symbol is developed as token combinations, then the relationship between each motion pattern starts to be constructed. In the final level, logical relationship between symbol combined with the physical relationship between

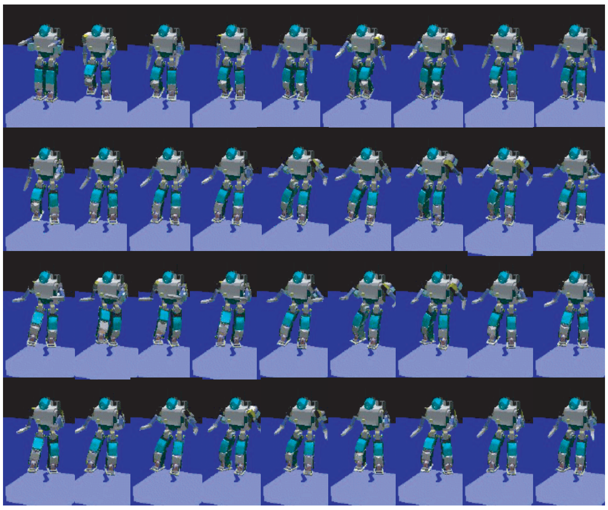

Figure 13: A result of motion pattern generation : kick after walk

motion pattern. Our approach follows the development model. At the present moment, our method achieved the transitional level and is going to achieve the final symbolic level.

Following effects are expected by realization of mutual development model between motion patterns and symbols;

1. **Motion learning by imitation** : It is possible for a humanoid to acquire target motion patterns by observation even if the body condition of learner differs from the one of teacher based on the transmission of symbol representation.

2. **Motion learning by instructions** : Design of humanoid motion has complexity because of the large number of degree of freedoms. Teacher's symbolic instruction helps the development of motion patterns.

3. **Constructive understanding of high level brain function** : This research aims the connection between motion patterns and symbols from engineering point of view. Additionally, brain science and psychology also have close relationships between language and behavior. Ripple effects from engineering will reach to the interdisciplinary research, that is combination of brain science, neuro science, psychology.

References

[1] V. Gallese and A. Goldman: Mirror neurons and the simulation theory of mind-reading, *Trends in Cognitive Sciences*, Vol. 2, No. 12, pp. 493–501, 1998.

[2] Merlin Donald: *Origins of the Modern Mind*, Harvard University Press, Cambridge, 1991.

[3] Terrence W. Deacon: *The symbolic species*, W.W. Norton & Company. Inc., 1997.

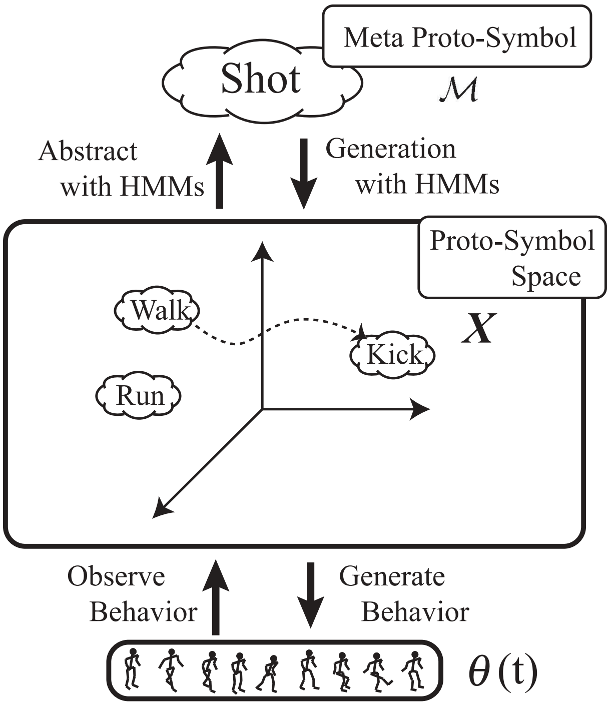

Figure 14: An outline of hierarchical mimesis model

[4] B.H. Juang and L.R. Rabiner: A probabilistic distance measure for hidden markov models, In *AT&T Technical Journal*, Vol. 64, No.2, pp. 391–408, 1985.

[5] P.K. Pook and D.H. Ballard: Recognizing teleoperated manipulations, In *IEEE International Conference on Robotics and Automation*, pp. 578–585, 1993.

[6] Tetsunari Inamura, Hiroaki Tanie, and Yoshihiko Nakamura: From stochastic motion generation and recognition to geometric symbol development and manipulation, In *International Conference on Humanoid Robots*, 2003.

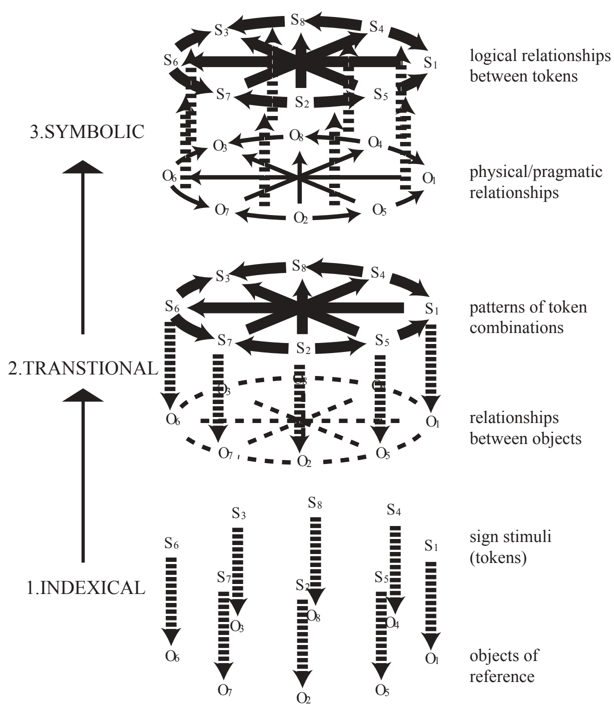

Figure 15: Development model of symbol proposed by Deacon