2009 IEEE International Conference on Robotics and Automation
Kobe International Conference Center
Kobe, Japan, May 12-17, 2009

Statistically Integrated Semiotics that Enables Mutual Inference Between Linguistic and Behavioral Symbols for Humanoid Robots

Wataru Takano and Yoshihiko Nakamura

***Abstract*— This paper describes the linguistic model based on symbolization of motion patterns for humanoid robots. The model consists of two kinds of stochastic models : the motion language model and the natural language model. The motion language model stochastically connects the symbols of motion patterns to the morpheme words through the latent states which represent the underlying linguistic structure such as semantic contents. The natural language model represents the dynamics of the word classes. The motion language model and the natural language model correspond to semantics and syntax respectively. The integration of the motion language model and the natural language model allows robots not only to linguistically interpret the motion patterns as sentences but also to generate the motions from the sentences. The two kinds of linguistic processes of the interpretation and the generation can be obtained by solving search problems: search for a sequence of morpheme words and a symbol of motion pattern. The proposed approach to interpretation of motion patterns as sentences and generation of motion patterns from the sentences through integration of the motion language model and the natural language model is validated by the experiment on the human behavioral data.**

I. INTRODUCTION

Language is crucial for intelligent processing of human beings. Language allows humans to memorize a lot of objects as symbols, to infer symbolically and to verbally communicate with one another. Deacon suggest a hypothesis that evolution of symbolic process is achieved hierarchically through icon, index and symbol [1]. The icon is a reference to an object under physical similarity. The index is a reference based on physical and temporal relations among the icons. The symbol is a more abstract reference to an object based on connection to other symbols. The hierarchical evolution of symbol enables humans to develop intelligence. In neuroscience, Rizzolatti discovered mirror neurons in primate's brain [2]. The mirror neurons active not only when the primate observes demonstrator's behavior but also when the primate performs the same kind of behavior. The loop of recognition and generation of motion in the mirror neurons contributes to communication as a mind-reading system [3]. Moreover, the mirror neurons, which are located near Broca's area, are related to symbolization, recognition and generation of motion patterns.

In robotics, some frameworks for symbolization of robot's motion patterns have been proposed, such as Recurrent Neural Network with Parametric Bias (RNNPB) [4], MOdule Selection And Identification for Control (MOSAIC) [5], and Hidden Markov Model approach (HMM) [6][7]. Additionally a model of nonverbal communication between a humanoid robot and its partner based on the symbolization of motion patterns is presented [8]. Although the communication model consists of two hierarchies of symbolic representations of motion patterns and behavioral interaction patterns, the communication is not based on linguistic processing. Research of integration of symbols of motion pattern and verb words has been made, where the symbols of motion patterns are stochastically associated with the verbs [9]. However, this framework deals with only verb words without other word classes such as noun or particle. Bidirectional processing of linguistic interpretation of motion patterns as sentences and understanding of sentences through generation or simulation of motion patterns from sentences is significant for robots to logically reason or verbally communicate with humans.

This paper describes integrating symbols of motion patterns with natural language processing. The proposed framework consists of motion language model and natural language model. In the motion language model, morpheme words are stochastically associated with symbols of motion patterns through latent variables. The latent variables represent underlying linguistic features which are unobservable, such as semantic contents. The natural language model represents the dynamics of morpheme words through latent variables, which correspond to unobservable word classes. The motion language model and the natural language model work as semantic graph and syntax graph respectively. The integration of the motion language and the natural language makes it possible for robots not only to interpret motion patterns as sentences (sequences of morpheme words) but also to generate motion patterns corresponding to sentences by solving the graphical search problems through the motion language model and the natural language model.

Section II presents motion language model and natural language model. Section III describes integration methodology which allows not only interpretation of observation as sentences but also generation of motion corresponding to input sentences. Experimental results on captured motion data was presented in section IV. And finally conclusion follows in section V.

II. MOTION LANGUAGE MODEL AND NATURAL LANGUAGE MODEL

A. Motion Language Model

Although textual or phonetical modality included in language can be measured, some of the linguistic modalities which underlie the linguistic structure cannot be observed.

W. Takano and Y. Nakamura are with Mechano-Informatics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan, {takano,nakamura} @ynl.t.u-tokyo.ac.jp

978-1-4244-2789-5/09/$25.00 ©2009 IEEE

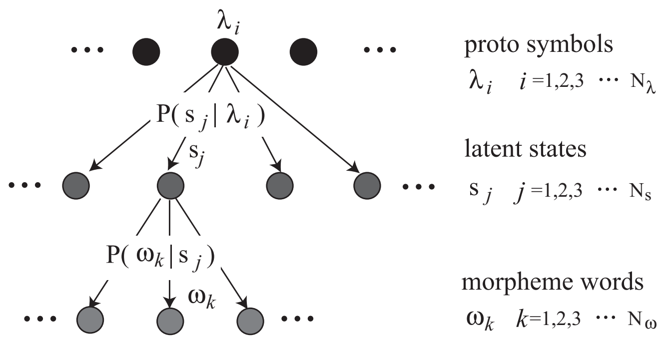

Fig. 1. The motion language model represent the stochastic association of Morpheme words with proto symbols via latent states. The motion language is defined by two kinds of parameters : probability that a morpheme word is generated by a latent variable and probability that a latent variable is generated by a proto symbol.

In this paper, we propose a motion language model stochastically connecting symbols of motion patterns to morpheme words through latent variables, which are suitable to represent unobservable states. Fig.1 illustrates the motion language model.

The motion language model consists of three kinds of nodes : symbol of motion pattern, latent variable and morpheme word. The symbol of motion pattern is represented by an HMM which learns spatial and temporal behavioral pattern [10]. The connections between the symbol of motion pattern and the latent variable and between the latent variable and the morpheme word are expressed by associative probabilities : $P(s|\lambda)$ and $P(\omega|s)$. Note that λ, s and ω are the symbol of motion pattern, the latent variable and morpheme word respectively, and that $P(s|\lambda)$ and $P(\omega|s)$ represent the probability that the symbol of motion pattern λ generates the latent variable s and the probability that the latent variable s generates the morpheme word ω.

The stochastic parameters of the motion language model $P(s|\lambda)$ and $P(\omega|s)$ are optimized by EM (Expectation Maximization) algorithm, which alternately processes two steps : Expectation step (E-step) and Maximization step (M-step). Training pairs of symbol of motion pattern and a sentence (a sequence of morpheme words) are given. The training pair is described by $\left\{\lambda^k; \omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k k = 1, 2, 3, \cdots, N\right\}$. Note that N is the number of training pairs and that n_k is the number of the morpheme words composing the k-th sentence. Both of the symbol of motion pattern λ^k and the sentence $(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k)$ represent k-th motion pattern.

E-step calculates distributions of the latent variables based on model parameters estimated in previous M-step. The distributions of the latent variables are provided as follows.

$$P(s|\lambda^k, \omega_i^k) = \frac{P(\omega_i^k|s, \lambda^k, \theta)P(s|\lambda^k, \theta)}{\sum_{j=1}^{N_s} P(\omega_i^k|s_j, \lambda^k, \theta)P(s_j|\lambda^k, \theta)} \tag{1}$$

where θ is a set of the previously estimated model parameters $P(s|\lambda)$ and $P(\omega|s)$.

M-step estimates the model parameters such that summation of expectation of log-likelihood that the symbol of motion pattern λ^k generates the sentence $(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k)$ is maximized. The summation of the expectation of the log-likelihood Φ is described by the following equation.

$$\Phi = \sum_{k=1}^{N} \log P(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k|\lambda^k) \tag{2}$$

$$P(\omega_1^k, \omega_2^k, \cdots, \omega_{n_k}^k|\lambda^k) = \prod_{i=1}^{n_k} P(\omega_i^k|\lambda_k) \tag{3}$$

$$P(\omega_i^k|\lambda^k) = \sum_{j=1}^{N_s} P(\omega_i^k|s_j)P(s_j|\lambda^k) \tag{4}$$

where we uses conditional independence assumption expressed by Equation 3. The probability that a symbol of motion pattern generates a morpheme word can be rewritten as Equation 4. The estimates of the new model parameters are as follows.

$$P(s|\lambda) = \frac{\sum_{k=1}^{N}\sum_{i=1}^{n_k} \delta(\lambda, \lambda^k)P(s|\lambda^k, \omega_i^k)}{\sum_{j=1}^{N_s}\sum_{k=1}^{N}\sum_{i=1}^{n_k} \delta(\lambda, \lambda^k)P(s_j|\lambda^k, \omega_i^k)} \tag{5}$$

$$P(\omega|s) = \frac{\sum_{k=1}^{N}\sum_{i=1}^{n_k} \delta(\omega, \omega_i^k)P(s|\lambda^k, \omega_i^k)}{\sum_{j=1}^{N_\omega}\sum_{k=1}^{N}\sum_{i=1}^{n_k} \delta(\omega_j, \omega_i^k)P(s|\lambda^k, \omega_i^k)} \tag{6}$$

where $\delta(\lambda_i, \lambda_j)$ and $\delta(\omega_i, \omega_j)$ are Kronecker deltas. $\delta(\lambda_i, \lambda_j)$ and $\delta(\omega_i, \omega_j)$ become 1 if i is equal to j. Otherwise, they become 0 respectively.

By iteratively computing the distributions of the latent variables and the estimates of model parameters by using Equations 1, 5 and 6, we can derives the appropriate motion language model.

B. Natural Language Model

Various kinds of language models have been proposed in a community of natural language processing. Especially, stochastic models are advantageous such as CRF (Conditional Random Fields) [11] or HMM [12] since the linguistic model is required to deal with a lot of words. In this paper we construct a natural language model by using HMM. Fig.2 illustrates the natural language model, where each node corresponds to a word class such as noun, verb, particle and so on. The node stochastically generates words that are classified to the node and the dynamics of the word classes are expressed by the stochastic transitions among the nodes. The natural language model is defined by a set of parameters: initial node distributions $\{\pi_i | i = 1, , 2, 3, \cdots, N_c\}$ that initial morpheme words are classified as the word class c_i, the transition probabilities $\{P(c_i|c_j) | i, j = 1, 2, 3, \cdots, N_c\}$ that the node c_i follows the node c_j, and the output probabilities $\{P(\omega_k|c_i) | i = 1, 2, 3, \cdots, N_c, k = 1, 2, 3, \cdots, N_\omega\}$ that

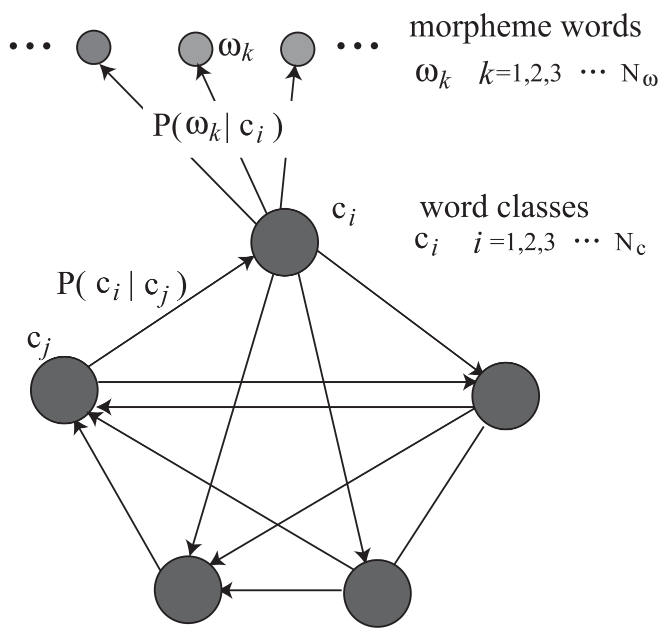

Fig. 2. Natural language model represents the dynamics of word classes by Hidden Markov Models. The node c_i corresponds to the word class. Transition from the node c_j to the node c_i is implemented with probability $P(c_i|c_j)$. The morpheme word ω_k is generated by the node c_i with conditional probability $P(\omega_k|c_i)$.

the node c_i generates the morpheme word ω_k, where N_c indicates the number of nodes in the natural language model.

The optimized model parameters can be derived from following equations.

$$\pi_i = \frac{N(\pi_i)}{\sum_{j=1}^{N_c} N(\pi_j)} \tag{7}$$

$$P(c_i|c_j) = \frac{N(c_i, c_j)}{\sum_{i=1}^{N_c} N(c_i, c_j)} \tag{8}$$

$$P(\omega_k|c_i) = \frac{N(\omega_k, c_i)}{\sum_{k=1}^{N_\omega} N(\omega_k, c_i)} \tag{9}$$

where $N(\pi_i)$ is the number of times that initial words are classified into the word class corresponding to the node c_i, $N(c_i, c_j)$ is the number of times that the word class c_i follows the word class c_j, $N(\omega_k, c_i)$ is the number of times that the word ω_k is categorized as the word class c_i in a set of training sentences.

III. Integration of the Motion Language Model and the Natural Language Model

The integration of the motion language model and the natural language model is required for robots to interpret not only their own behaviors but also their partner's ones as sentences linguistically based on the symbolic representations of motion primitives. Especially this paper describes bidirectional processes between the symbols of motion patterns and the sequences of morpheme words as illustrated by Fig.3. One process is interpretation of a motion pattern as a sentence and the other is generation or simulation of a motion pattern corresponding to a given sentence. These two computational processes are described by the stochastic searching problems.

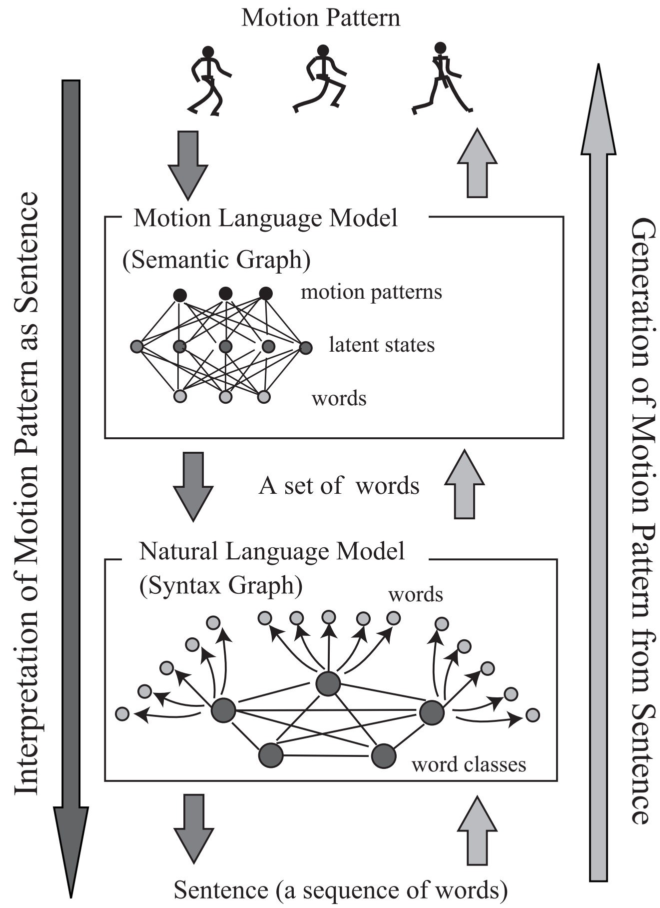

Fig. 3. Overview of integration of a motion language model with a natural language model. The motion language model represents relationship among proto symbols and morpheme words via latent variables as a graph structure, where nodes on 1st, 2nd and 3rd layer indicate the proto symbols, the morpheme words and the latent variables respectively. The natural language model represents the dynamics of language which means the order of words in sentences. The motion language model and the natural language model are equivalent to semantics and syntax. By integrating two functions, linguistic processing for robots can be realized.

A. Interpretation of Motion Pattern as Sentence

The interpretation of motion pattern as sentence can be made through recognition of a motion pattern as a symbol and form of a sentence corresponding to the symbol using the motion language model and the natural language model.

A motion pattern can be recognized as a symbol with the largest likelihood that the motion pattern is generated by a symbol. The motion recognition can be processed as follows.

$$\lambda^o = \arg \max_{\lambda_i : i=1,2,\cdot,N_\lambda} P(\boldsymbol{O}|\lambda_i) \tag{10}$$

where $P(\boldsymbol{O}|\lambda_i)$ is the likelihood that the motion pattern $\boldsymbol{O}$ is generated by the proto symbol λ_i, and the symbol λ^o is the result of motion recognition. Note that the motion pattern $\boldsymbol{O}$ is represented by temporal-spatial data such as a sequence of joint angles.

Forming a sentence corresponding to the symbol of motion pattern becomes searching a sequence of morpheme words which is the most likely generated by the symbol of motion pattern. The search problem can be described as follows.

$$\begin{aligned}
\boldsymbol{\omega^o} &= \arg\max_{\forall\boldsymbol{\omega}} P(\boldsymbol{\omega}|\lambda^o, L) && (11)\\
&= \arg\max_{\forall\boldsymbol{\omega}} \frac{P(\boldsymbol{\omega}|L)P(\lambda^o|\boldsymbol{\omega}, L)}{P(\lambda^o|L)} && (12)\\
&= \arg\max_{\forall\boldsymbol{\omega}} P(\boldsymbol{\omega}|L)P(\lambda^o|\boldsymbol{\omega}) && (13)\\
&= \arg\max_{\forall\boldsymbol{\omega}} \frac{P(\boldsymbol{\omega}|L)P(\boldsymbol{\omega}|\lambda^o)P(\lambda^o)}{P(\boldsymbol{\omega})} && (14)\\
&= \arg\max_{\forall\boldsymbol{\omega}} P(\boldsymbol{\omega}|\lambda^o)P(\boldsymbol{\omega}|L) && (15)
\end{aligned}$$

where L is the natual language model and $\boldsymbol{\omega}$ is a sequence of morpheme words; $\boldsymbol{\omega} = \left\{\omega_1^*, \omega_2^*, \cdots, \omega_{n_*}^*\right\}$. We can obtain the transformation from Equation 11 to Equation 12 and from Equation 13 to Equation 14 by using Bayes rule. Equation 12 can be transformed to Equation 13 since $P(\lambda^o|L)$ does not depend on a sequence of morpheme words $\boldsymbol{\omega}$. Equation 15 can be derived from assumption that the prior probability distributions $P(\lambda^o)$ and $P(\boldsymbol{\omega})$ are equiprobable. $P(\boldsymbol{\omega}|\lambda^o)$ and $P(\boldsymbol{\omega}|L)$ can be computed by using the motion language model and the natural language model. The search problem expressed by Equation 15 is too complicated since the search space is infinite. However the equation 15 can be approximated as follows with help of log-likelihood and Viterbi algorithm so that the search problem can be solved efficiently.

$$\boldsymbol{\omega^o} \approx \arg\max_{\forall\boldsymbol{\omega}} \left[\log P(\boldsymbol{\omega}|\lambda^o) + \log P(\boldsymbol{\omega}|\boldsymbol{c^\omega}, L)\right] \quad (16)$$

$$P(\boldsymbol{\omega}|\lambda^o) \leq \sum_{i=1}^{k} \log P(\omega_i^*|\lambda^o) \quad (17)$$

$$\log P(\boldsymbol{\omega}|\boldsymbol{c^\omega}, L) \leq \log \pi_{c_1^\omega} + \sum_{i=2}^{k} \log P(c_i^\omega|c_{i-1}^\omega) + \sum_{i=1}^{k} \log P(\omega_i^*|c_i^\omega) \quad (18)$$

where $\boldsymbol{c^\omega}$ is the Viterbi path expressed by $\left\{c_1^\omega, c_2^\omega, \cdots, c_{n_*}^\omega\right\}$, which is the most likely sequence of the nodes in the natural language model for given sequence of the morpheme words. This path can be found by the Viterbi algorithm. The searching problem expressed by Equation 16 can be solved efficiently by A^* search algorithm by using Equations 17 and 18 [13]. When the morpheme words of $\{\omega_1^*, \omega_2^*, \cdots, \omega_k^*\}$ is determined and the morpheme words of $\left\{\omega_{k+1}^*, \omega_{k+2}^*, \cdots, \omega_{n_*}^*\right\}$ is not yet, the evaluation value for the sequence of morpheme words $\boldsymbol{\omega}$ described by Equation 16 can be overestimated as sum of Equation 17 and Equation 18 by using the admissible heuristic estimate since probabilities $P(\omega_i^*|\lambda^o)$, $P(c_i^\omega|c_{i-1}^\omega)$ and $P(\omega_i^*|c_i^\omega)$ are less than or equal to 1. Employing the overestimate described above allows the efficient search. Note that the search using the overestimate is guaranteed to reach an optimal solution.

B. Generation of Motion Pattern from Sentence

A motion pattern is generated from a given sentence through searching a symbol of motion pattern corresponding to the sentence and regenerating temporal-spatial data from the symbol. The search problem for the symbol of motion pattern corresponding to the sentence can be described as follows.

$$\begin{aligned}
\lambda^o &= \arg \max_{\lambda_i : i=1,2,\cdots,N_\lambda} P(\lambda_i|\boldsymbol{\omega}) && (19)\\
&= \arg \max_{\lambda_i : i=1,2,\cdots,N_\lambda} \frac{P(\lambda_i, \boldsymbol{\omega})}{P(\boldsymbol{\omega})} && (20)\\
&= \arg \max_{\lambda_i : i=1,2,\cdots,N_\lambda} P(\lambda_i)P(\boldsymbol{\omega}|\lambda_i) && (21)\\
&= \arg \max_{\lambda_i : i=1,2,\cdots,N_\lambda} P(\boldsymbol{\omega}|\lambda_i) && (22)\\
&= \arg \max_{\lambda_i : i=1,2,\cdots,N_\lambda} \sum_{j=1}^{n_*} \log P(\omega_j^*|\lambda_i) && (23)
\end{aligned}$$

where $\boldsymbol{\omega}$ is a input sentence $\left\{\omega_1^*, \omega_2^*, \cdots, \omega_{n_*}^*\right\}$ and λ^o is the symbol of motion pattern corresponding to the sentence. Equation 19 is transformed to Equation 20 by using Bayes rule. Since $P(\boldsymbol{\omega})$ does not depend on symbols of motion patterns, Equation 21 can be derived with the help of Bayes rule. The transformation from Equation 21 to Equation 22 can be derived based on assuming that prior probability distributions $P(\lambda)$ are equiprobable. The search problem results in Equation 23 with the help of log likelihood and Equation 3.

Equation 23 can be solved efficiently through A^* search algorithm since the right side in Equation 23 can be overestimate as follows.

$$\sum_{j=1}^{n_*} \log P(\omega_j^*)|\lambda_i) \leq \sum_{j=1}^{k} \log P(\omega_j^*)|\lambda_i) \quad (24)$$

where $k \leq n_*$. The node with the largest evaluation described by Equation 24 is expanded incrementally. Note that the expansion is the process where the node is re-evaluated using the sequence of morpheme words $\left\{\omega_1^*, \omega_2^*, \cdots, \omega_{k+1}^*\right\}$. The procedure of the search is depicted by Fig.4.

IV. Experiments

The integration of the motion language model and the natural language model was tested on human motion data obtained through a optical motion capture system. The motion capture system measures the positions of 34 markers attached to a performer. The sequences of marker positions are converted to the sequences of joint angles by inverse kinematics computation based on a humanoid robot with 20 degrees of freedom. The human motion data set contains 10 kinds of motion patterns related to baseball : “running”, “jumping”, “swinging a bat”, “throwing a ball”, “sliding”, “catching a ball”, “diving”, “standing up”, “shaking a hand”, and “crouching”. These motion data are used as training data for HMMs. 10 symbols of motion patterns $N_S = 10$ are

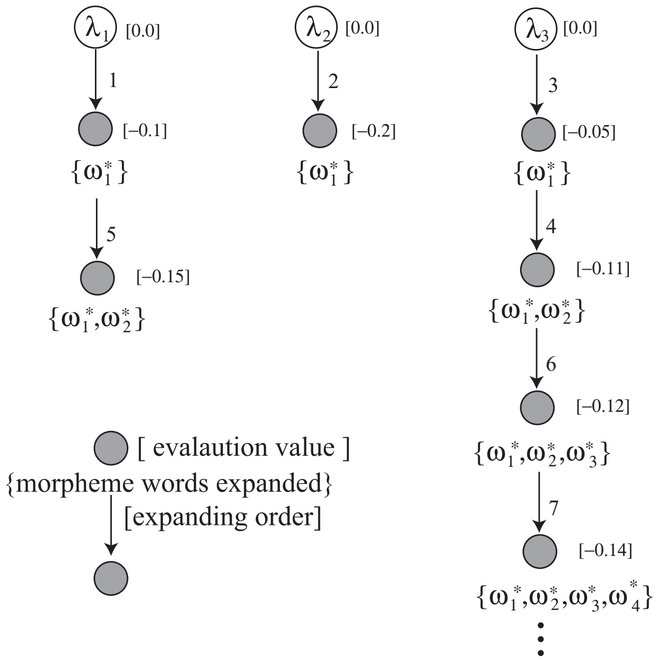

Fig. 4. This figure shows an overview of the search algorithm for the symbol of motion pattern corresponding to input sentence in the case that the number of the symbols is 3. The node with the largest evaluation is expanded incrementally. The expansion is the procedure of adding a morpheme word to the node and re-evaluating the node.

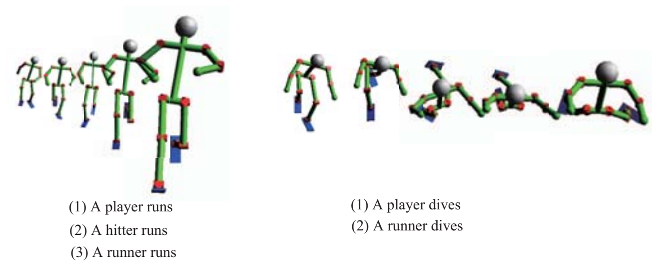

Fig. 5. A sentence corresponding to each motion pattern is generated by the integrated model of the motion language model and the natural language model. The sentences corresponding to the motion pattern are displayed in order of the likelihood that the sentence is generated from the symbol of the motion pattern through the motion language model and the natural language model. The three sentences : "*start* a player runs *end*", "*start* a hitter runs *end*" and "*start* a runner runs *end*" represent the motion pattern of "running". The two sentences : "*start* a player dives *end*" and "*start* a runner runs *end*" represent the motion pattern of "diving"

TABLE I

EVALUATION OF LINGUISTIC EVALUATION

motion pattern	evaluation Ψ
running	0.99
jumping	0.99
swinging a bat	0.92
throw a ball	0.92
sliding	0.99
catching a ball	0.99
diving	0.99
standing up	0.99
shaking a hand	0.99
crouching	0.89

acquired by optimizing the parameters of HMMs with the motion data.

The human motion data are expressed not only by the symbols of motion patterns but also by the Japanese sentences (the sequences of morpheme words). For examples, the motion data of "crouching" is given four kinds of sentences: "*start* a player crouches *end*)", "*start* a pitcher crouches *end*", "*start* a pitcher crouches on the mound *end*" and "*start* on the mound, a pitcher crouches *end*)". 23 ($N = 23$) training pairs of the motion pattern and the sentence are used to obtain the motion language model, which consists of 50 latent variables ($N_S = 50$) and 24 morpheme words ($N_\omega = 24$).

The natural language model is also optimized by using the same sentences that the motion language model uses for learning. The natural language model contains 5 nodes. The morpheme words are classified as one of five word classes : "noun", "verb", "particle", "start-class" and "end-class". Note that "start-class" and "end-class" indicate the beginning and the end of the sentences.

A. Experimental Result of Interpreting Motion Pattern as Sentence

By using the motion language model and the natural language model, the proposed approach to interpretation of motion patterns as sentences was tested. The captured human motion data is recognized as the symbol of motion pattern. The symbol generates sentences through the motion language model and the natural language model. Fig.5 shows the experimental results of interpreting motion patterns as sentences. In Fig.5, the motion pattern of "running" can be interpreted as appropriate sentences which express that a player, hitter or runner runs. Additionally, the motion pattern of "diving" can be also correctly interpreted as sentences : "a player dives" and "a runner dives".

The evaluation experiment for the proposed framework of the linguistic interpretation was made. The framework is evaluated based on the likelihood that correct sentences corresponding to the input motion pattern are generated. The evaluation is calculated as follows. Some sentences corresponding to the input motion pattern are found as solutions. If the found solution of a sentence is included in the set of training pairs of the input motion pattern, the solution is correct. Otherwise the solution is wrong. Moreover likelihood that the solution is generated is calculated by using Equations 3 4. Therefore, the evaluation can be defined as follows

$$\Psi = \frac{\displaystyle\sum_{\boldsymbol{\omega_i} \in S_{correct}: i=1,2,\cdots,n_c} P(\boldsymbol{\omega_i}|\lambda^o)}{\displaystyle\sum_{\boldsymbol{\omega_i}: i=1,2,\cdots,n_c} P(\boldsymbol{\omega_i}|\lambda^o)} \tag{25}$$

where Ψ is the evalaution of the linguistic interpretation, λ^o is the symbol of input motion pattern, $\boldsymbol{\omega_i}$ is the found solution of a sentence, n_c is the number of the solution and $S_{correct}$ is the set of sentences included in the training pairs corresponding to the input motion pattern. Although

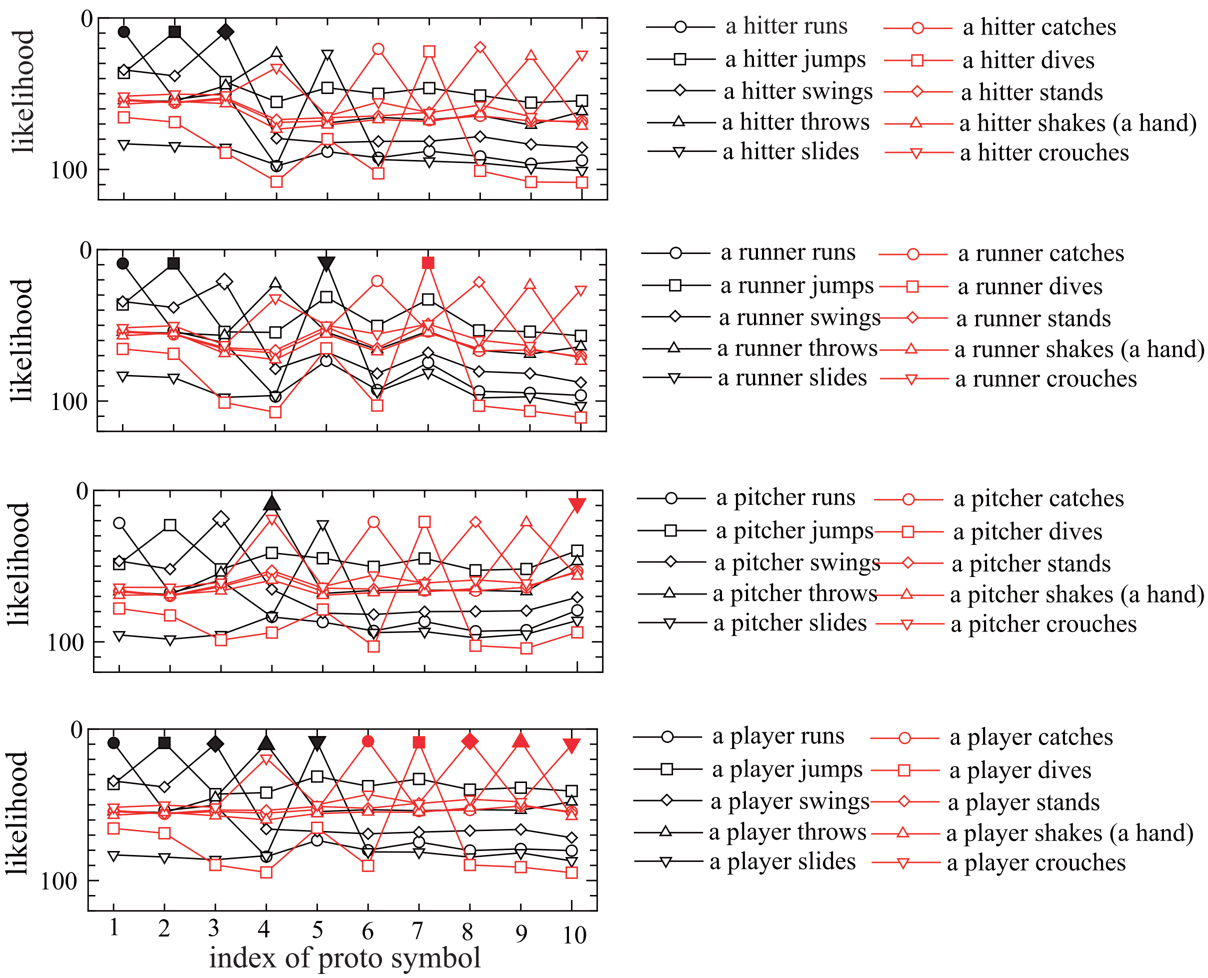

Fig. 6. These figures show the likelihood that each proto symbol is generated by simple sentence consisting of "subject" and "verb". Proto symbols represent motion patterns of "run", "jump", "swing", "throw", "slide", "catch", "dive", "stand", "shake a hand" and "crouch" in ascending order. For example, the proto symbol of run is associated with a sentence that a hitter runs since the likelihood that this proto symbol is generated by the sentence is the largest. Therefore proto symbols are correctly associated with all of the sentences. Especially appropriate association can be achieved even if the untrained sentences are input. Note that filled markers denote the trained pairs of a proto symbol and a sentence.

the length of sentence is not fixed and the search space is infinite, the number of the solution is set to 10 $n_C = 10$ in this evaluation experiment.

Table I shows the evaluation of the method to interpret the motion pattern as sentences. The motion patterns which have sentence structures of noun-particle-verb can be interpreted as correct sentences with probability of 0.99. The motion patterns which have sentence structures of noun-particle-object-particle-verb can also be interpreted as correct sentences with large probability of 0.92. The motion pattern of "crouching" is expressed by sentences with both of above two kinds linguistic topologies. This motion pattern is less likely interpreted as correct sentences than any other motion pattern. However correct interpretation can be made with large probability. The evaluation experiment demonstrates the validity of proposed framework of interpreting the motion pattern as the sentence through the motion language model and the natural language model.

B. Experimental Result of Generation of Motion Pattern from Sentence

The algorithm to generate the motion pattern from the sentence was tested. Fig.6 shows the likelihood that each symbol of motion pattern is generated from every sentence with a sentence pattern of subject and verb. Proto symbols represent "run", "jump", "swing", "throw", "slide", "catch", "dive", "shake a hand" and "crouch" in ascending order. The sentences including verb of "run" are the most likely to generate the proto symbol of "run". Other sentences also generate correct proto symbols. Especially, filled marks correspond to pairs of proto symbol and sentence, which are used in learning the motion language model. These pairs have larger likelihood. Other untrained pairs, such as of 4-th proto symbol and "a hitter throw", have smaller likelihood than the trained pairs. However, correct proto symbols are generated by input sentences. Therefore, This experiment validates that the proposed framework can be used not only for trained pairs of proto symbol and sentence but also untrained ones.

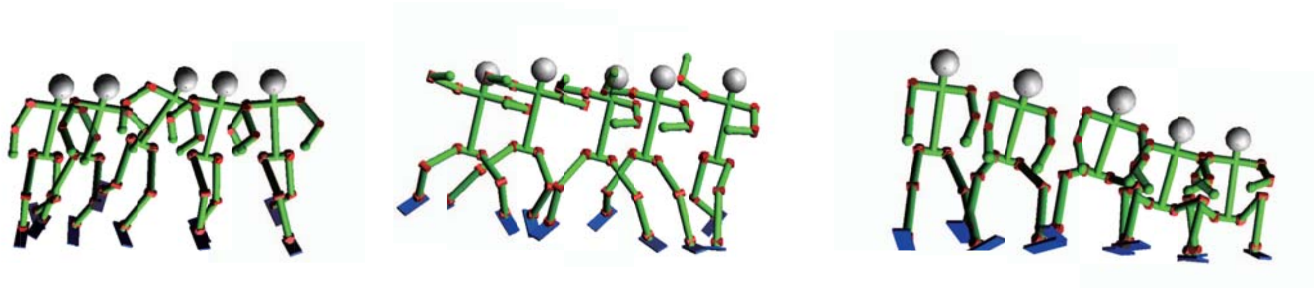

Fig. 7. The three kinds of motion data are generated or simulated from the input sentences : "*start* A player runs *end*", "*start* A runner throws a ball *end*" and "*start* A pitcher crouches on the mound *end*". The adequate motion data are generated in response to the sentences through the motion language model.

Fig.7 shows the motion data generated from the three kinds of sentences, "subject and verb", "subject, verb and object", and "subject, verb and adverb". The generated motion pattern can be categorized into the motion patterns of "running","throwing" and "crouching" respectively. These proto symbols appropriately correspond to input sentences. Therefore, this experiment demonstrates the validity of the framework to generate the motion patterns from not only simple sentence described in "subject and verb" but also a little complicated sentences in "subject, verb and object" and "subject verb and adverb".

V. Conclusion

The contributions of this paper are summarized as follows:

1) This paper describes the motion language model which connects morpheme words to motion patterns via latent variables stochastically. The motion language model is defined by two kinds of parameters : the probability that the motion pattern generates the latent variable and the probability that the latent variable generates the morpheme word. The latent variables are expected to represent unobservable data and reveal the underlying linguistic structure.
2) We integrates the motion language model with the national language model. The novel approach to interpretation of motion patterns as sentences through the integration of the motion language model and the natural language model is developed.
3) The method to generate the motion from the sentence is also proposed by using the motion language model.
4) The computational algorithm for integration of the motion language model and natural language model can be successfully implemented. The implementation demonstrates that semantically and syntactically appropriate sentences can be associated with motion patterns. Additionally the implementation validates the generation of the motion pattern from the sentence.

Acknowledgment

This research was supported by Category S of Grant-in-Aid for Scientific Research (20220001), Japan Society for the Promotion of Science.

References

[1] T. Deacon, "Symbolic species : The co-evolution of language and the brain." W.W.Norton and Company Inc, 1997.
[2] G. Rizzolatti, L. Fogassi, and V. Gallese, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nature Reviews*, pp. 661–670, 2001.
[3] V. Gallese and A. Goldman, "Mirror neuron and the simulation theory of mind-reading," *Trends in Cognitive Sciences*, vol. 2, no. 12, pp. 493–501, 1998.
[4] J. Tani and M. Ito, "Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment," *IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans*, vol. 33, no. 4, pp. 481–488, 2003.
[5] M. Haruno, D. Wolpert, and M. Kawato, "Mosaic model for sensorimotor learning and control," *Neural Computation*, vol. 13, pp. 2201–2220, 2001.
[6] A. Billard and R. Siegwart, "Robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 47, pp. 65–67, 2004.
[7] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *International Journal of Robotics Research*, vol. 23, no. 4, pp. 363–377, 2004.
[8] W. Takano, K. Yamane, T. Sugihara, K. Yamamoto, and Y. Nakamura, "Primitive communication based on motion recognition and generation with hierarchical mimesis model," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2006, pp. 3602–2609.
[9] W. Takano, K. Yamane, and Y. Nakamura, "Capture database through symbolization, recognition and generation of motion patterns," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2007, pp. 3092–3097.
[10] W. Takano and Y. Nakamura, "Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 425–431.
[11] T. Kudo, K. Yamamoto, and Y. Matsumoto, "Applying conditional random fields to japanese morphological analysis," in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 2004, pp. 230–237.
[12] K. Takeuchi and Y. Matsumoto, "Hmm parameter learning for japanese morphological analyzer," in *Proceedings of the 10th Pacific Asia Conference on Language, Information and Computation*, 1995, pp. 163–172.
[13] R. Dechter and J. Pearl, "Generalized best-first search strategies and the optimality of A*," *Journal of the ACM*, vol. 32, no. 3, pp. 505–536, 1985.