

Towards Interactive Physical Robotic Assistance: Parameterizing Motion Primitives Through Natural Language

José Ramón Medina¹, Michael Shelley¹, Dongheui Lee¹, Wataru Takano² and Sandra Hirche¹

Abstract—Natural language interaction between humans and robots is a very challenging topic, especially when it refers to motion descriptions in a certain environment. This problem is particularly relevant during physical human-robot interaction, e.g. in cooperative transportation tasks, where the partners’ physical coupling requires an agreement on the way to follow. Understanding in depth the link between sentences, words, environmental properties and motions can deeply enhance the interaction between humans and robots. In this work, we propose a novel approach for learning relations and dependencies between motion, natural language and environmental properties using parameterized left-to-right time-based Hidden Markov Models. A natural language model represents the link between language and motion symbols while the HMMs parameterization corresponds to the explicit influence on motions of both words and environmental features. The proposed PHMM approach parameterizes the output and the transition probabilities using a non-linear dependency estimation. The method is validated by learning and generating navigation primitives in a 2 Degrees-Of-Freedom (DoF) virtual scenario.

I. INTRODUCTION

In order for robots to interact with humans intuitively they must understand their partners on different levels of communication - verbally and non-verbally. This is particularly important for the cooperation of robots with humans under uncertainty where interactive planning, decision making, and control play a crucial role as e.g. in [1]. Desirably a robot understands a verbal “command” by the human interaction partner and executes the corresponding motor action. As an example let us consider a human-robot joint transportation task. Due to the physical coupling, both partners must agree on the path to follow as they navigate through the environment. In case of disagreement, humans can easily communicate with each other using natural language in order to find a consensus and proceed with the task execution. Synthesizing similar interactive behavior for robotic partners is however a very challenging problem as the information exchange between partners usually refers to environmental properties.

Humans easily interact with each other using motions and gestures [2], [3]. Natural language is a unique symbolic communication system to humans that provides much richer descriptions even when referring to motions. More recently also the interaction through haptic signals as in physical human-robot cooperation is of increasing interest [1], [4].

One of the major challenges is the combination of natural language with motor action in terms of haptic primitives with consideration for environmental and task constraints. Understanding the link between signifiers and their corresponding significant is a very challenging topic [5], [6], especially considering motions [7]. This problem usually implies, on one side, modeling sentence structures and word relations in a natural language model and, on the other side, considering the association between sentences and motion symbols [8], [9]. Still, this representation does not express the quality of behaviors in an explicit way. Many adverbs and adverbials specify how to parametrize and/or change motions or haptic primitives. As an example take the sentence “Turn slowly” where “turn” stands for a motion symbol and “slowly” expresses the quality of the motion primitive. Extracting such word models can deepen robots’ comprehensive capabilities when using natural language and improve their inference performance. Furthermore, motions, as well as their natural language descriptions, are usually influenced by environmental characteristics and constraints. Considering these extrinsic dependencies in conjunction with qualitative natural language descriptions will significantly expand robots understanding capabilities and will potentially shrink the gap towards seamless HRI.

The representation of time-series is a well-know issue in the fields of learning by demonstration or speech recognition. Hidden Markov Models (HMMs) are widely used in the literature as a compact spatio-temporal representation of behaviors [10]. Adding explicit time information to the standard model, generalized motions can be represented by smooth trajectories using time-based HMMs (tHMMs) [11]. When behaviors are characterized by external variables, a natural extension of standard HMMs are Parametric HMMs (PHMMs), which include dependencies of the standard model parameters w.r.t external parameters [12]. PHMMs have been used to enrich motions representation by parameterizing its output densities [12] or by linearly interpolating exemplar parameters of the different states [13]. However, a full parameterization of all HMM parameters in a non-linear fashion is still an open issue.

In this paper we address the problem of learning and executing motion primitives including quality descriptors and environmental constraints. To this end we propose a novel approach for learning relations and dependencies between motion symbols, natural language and environmental properties using parameterized left-to-right tHMMs. A natural language model together with a motion language model infers the link between sentences and motion symbols while

¹J. R. Medina, M. Shelley, D. Lee and S. Hirche are with the Department of Electrical Engineering and Information Technology, Technische Universität München, D-80290 Munich, Germany. {medina, shelley, dhlee, hirche}@tum.de

²W. Takano is with Mechano-Informatics, The University of Tokyo, takano@ynl.t.u-tokyo.ac.jp

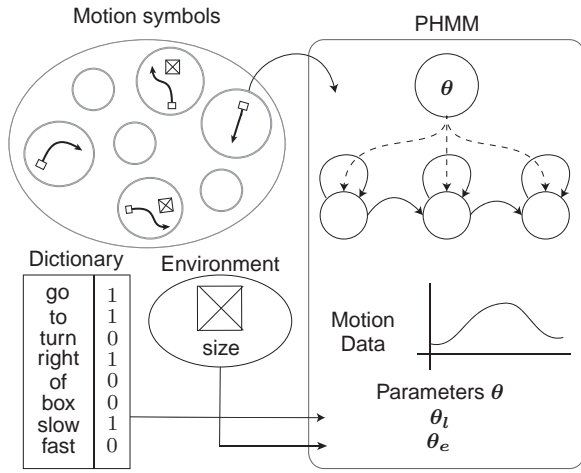


Fig. 1. PHMM representation. Each motion symbol is represented by a PHMM trained with parametric observations. The parameters are given by the words used in the description of the motion and the environmental properties.

the tHMMs parameterization represents the explicit influence on motions of both words and environmental properties. The proposed PHMM approach directly adapts both the output and the transition probabilities of tHMMs providing a full non-linear dependency estimation. The method is validated by learning and generating navigation primitives in a 2 DoF virtual scenario.

The remainder of this paper is structured as follows: in Section II the general approach is presented. The proposed PHMM learning method is explained in Section III and the natural language model used is described in Section IV. Experimental results and discussion are presented in Section V.

II. GENERAL ARCHITECTURE

The work presented in this paper consists of two problems: On one side, learning explicit dependencies between words, environmental properties and motion primitives and, on the other side, generating the expected motion given a natural language query and a specific environmental situation.

For the first problem, we consider a set of motion primitives where each motion is represented by a PHMM trained with a set of parametric motions $\mathcal{O}_\theta = \{\mathcal{O}, \theta\}$, consisting of the motion itself \mathcal{O} and its corresponding parameters θ . In this work we represent motion primitives in task space. Depending on the reference frame, motions representing the same primitive might vary significantly if represented in classical task coordinates. In order to avoid this we choose an invariant representation of motions [14]. Each invariant motion \mathcal{O} is associated with a set of parameters $\theta = \{\theta_l, \theta_e\}$ given by the word parameters θ_l and the environmental parameters θ_e . The word parameters are determined by the natural language description of the motion. In order to avoid computationally intractable parameter spaces, instead of considering directly each word, a syntactic analysis selects potential adverbials constructions and verbal forms, drastically reducing the possible parameter candidates. As understanding natural language syntax lies out of the scope of

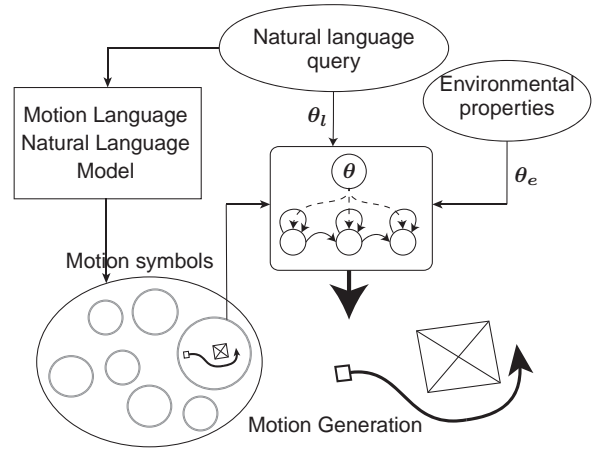


Fig. 2. Proposed architecture for motion generation. Given a query, the natural language model together with the motion language model extracts the most likely motion symbol from the database. With the parameters given by the words from the sentence in conjunction with the observed environmental properties, the most likely motion is generated.

this work, for simplicity of illustration we avoid this step and represent the word parameters as a vector with the size of the dictionary, where the words used in the motion description have value 1 and the others 0. The environmental parameters are a set of features of the environment such as an object size or its position. With this problem setting, shown in Fig. 1, the training process is carried out as described in Section III. Furthermore, in order to extract a generalized representation of how parameters affect motions, the set of primitives can be trained sharing the same parameterization function acquiring a *generalized parameter model*, as explained in Section III-B.

For the second problem, once the motion symbols are learned, a *motion generation* procedure produces the motion primitive given a natural language query and a specific environmental situation, as depicted in Fig. II. A natural language model together with a motion language model select the most likely motion symbol from the set of motion primitives, as described in Section IV. The words used in the query together with the perceived environmental features are then used to parameterize the selected motion symbol and generate the expected behavior, as explained in Section III-C.

III. PARAMETRIC TIME-BASED HIDDEN MARKOV MODELS

A standard left-to-right time-based HMM λ is composed of N states where each state has continuous observation densities represented by a Gaussian Mixture Model (GMM) with K components. λ is described by a set of parameters $\{\pi, \mathbf{a}, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$, where:

- $\pi = \{\pi_i\}$ is the initial state probabilities vector, for $1 \leq i \leq N$.
- $\mathbf{a} = \{a_{ij}\}$ is the transition probabilities matrix for $1 \leq i, j \leq N$.
- $\mathbf{w} = \{w_{ik}\}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_{ik}\}$, $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_{ik}\}$ are the weight, the mean and the covariance matrix of the GMMs for

both the spatial and *temporal* data for $1 \leq i \leq N$ and $1 \leq k \leq K$.

A parameterized observation $\mathbf{O}_\theta = \{\mathbf{O}, \theta\}$ is given by the observed spatial data sequence \mathbf{O} and its corresponding parameters θ . Note that the parameters are constant during the whole observation, i.e., a static environment is assumed.

The parametric tHMM modifies the standard output densities and the transition probabilities as a function of θ , i.e.

$$\begin{aligned} a_{ij}(\theta) &= \overline{a_{ij}} + f_{a_{ij}}(\theta) \\ \mu_{ik}(\theta) &= \overline{\mu_{ik}} + f_{\mu_{ik}}(\theta) \\ \Sigma_{ik}(\theta) &= \overline{\Sigma_{ik}} + f_{\Sigma_{ik}}(\theta), \end{aligned} \quad (1)$$

where $\overline{a_{ij}}$, $\overline{\mu_{ik}}$ and $\overline{\Sigma_{ik}}$ are the standard transition probabilities, means and covariances, calculated using the EM algorithm [15] using only the spatial data \mathbf{O} . Here we assume that the initial state probabilities are equal for all states and therefore no parameterization is needed, but a similar dependency could be easily added. The model's parametric dependency $\mathbf{f}(\theta) = \{f_{a_{ij}}(\theta), f_{\mu_{ik}}(\theta), f_{\Sigma_{ik}}(\theta)\}$ is estimated observing the variations of the standard tHMM parameters with respect to the external ones.

A. Parametric Dependency Learning

In order to extract a non-linear approximation of the parametric dependency, we model the joint probability density of $\mathbf{f}(\theta)$ and the external parameters, i.e. $P(\mathbf{f}(\theta), \theta)$, using a GMM. This probabilistic model expresses the probability distribution of the parameters θ in conjunction with the expected values of $\mathbf{f}(\theta)$. Given a set of D parametric observations ($d = 1 \dots D$), the joint distribution is estimated maximizing the likelihood of the GMM w.r.t the samples $\{\mathbf{f}(\theta_d), \theta_d\}$, where $\mathbf{f}(\theta_d)$ is given by

$$\begin{aligned} f_{a_{ij}}(\theta_d) &= a_{ij}(\theta_d) - \overline{a_{ij}} \\ f_{\mu_{ik}}(\theta_d) &= \mu_{ik}(\theta_d) - \overline{\mu_{ik}} \\ f_{\Sigma_{ik}}(\theta_d) &= \Sigma_{ik}(\theta_d) - \overline{\Sigma_{ik}}, \end{aligned} \quad (2)$$

and $a_{ij}(\theta_d)$, $\mu_{ik}(\theta_d)$ and $\Sigma_{ik}(\theta_d)$ are the transition probabilities, means and variances maximizing the likelihood only for the d -th observation.

A direct comparison of the standard model parameters with the observation specific ones requires an alignment of the states, i.e., a similar distribution of the responsibility of the states through all observations. In order to achieve this, we follow [13] initializing the values $a_{ij}(\theta_d)$, $\mu_{ik}(\theta_d)$ and $\Sigma_{ik}(\theta_d)$ with the overall ones $\overline{a_{ij}}$, $\overline{\mu_{ik}}$ and $\overline{\Sigma_{ik}}$ and then fixing the means after one iteration of the EM algorithm. This way, a similar distribution of the responsibility is extracted from the overall model validating the state comparisons from (2). The proposed approach is indeed similar to [13]. However, the inclusion of the GMM in order to model the parametric dependency provides better generalization properties due to its non-linearity and its regression capabilities in the whole parameter space.

Given the set of samples $\{\mathbf{f}(\theta_d), \theta_d\}$, the GMM is trained using the EM algorithm.

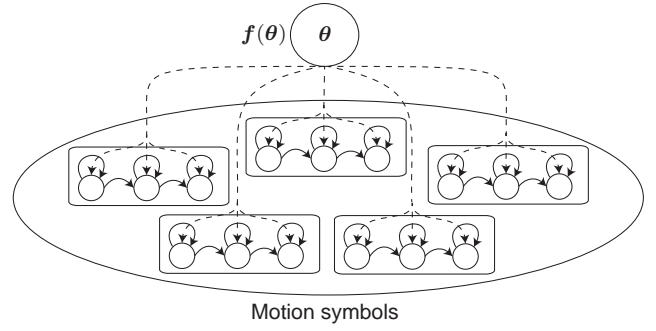


Fig. 3. Generalized Parameter model. The non-linear dependency $\mathbf{f}(\theta)$ is designed considering the parameterization of all possible motion symbols.

B. Generalized Parameter Model

While the influence of external parameters on a specific model is captured training a PHMM, a generalization of how those parameters affect motions can not be inferred with the parameterization of different primitives. In order to acquire a *generalized parameter model* the function $\mathbf{f}(\theta)$ must be learned maximizing the expected parameterization of all different motion primitives, as depicted in Fig. 3.

Given a set of M PHMMs ($m = 1 \dots M$) with D parametric observations ($d = 1 \dots D$) each, the generalized parameter model $\mathbf{f}(\theta)$ considers all $\{\mathbf{f}_m(\theta_d), \theta_d\}$ for $1 \leq m \leq M$ and $1 \leq d \leq D$, where $\mathbf{f}_m(\theta_d)$ is the expected value of the parametric function for the m -th PHMM and the d -th observation.

C. Inference and Motion Generation

Given a query with its corresponding parameters θ_q , generating the expected motion from a PHMM requires first the calculation of the expected values of $\mathbf{f}(\theta)$. Using Gaussian Mixture Regression (GMR) [16], we calculate the expected parameterization as the posterior mean estimate given by θ_q , i.e. $P(\mathbf{f}(\theta)|\theta_q)$. Applying then (1), the expected motion is generated calculating the generalized output of the resulting tHMM as explained in [11].

IV. MOTION AND NATURAL LANGUAGE MODEL

The connection between motions with natural language is achieved integrating two components: The motion language model and the natural language model. On one side, the natural language model represents stochastically all possible words sequences. On the other side, the motion language model is a stochastic model that captures the link between motion symbols and words. The combination of both methods provides a model that considers both the sentence structure constraints and the expected words and motion symbols links.

The natural language model stochastically represents sequences of words. A bigram model is used to represent links between words, modeling the expected sequences. Words are represented by nodes and transitions between two words, i.e. a word is followed by another one, as edges. This way, the observed sentence structures are expressed as transitions among words.

The motion language model represents the relations between motion symbols and words in a stochastic way. The model is structured in three different layers: motion symbols, latent states and words. Two different probabilities connect the motion symbols layer with the words layer. On one side the connection between motion symbols and latent states and on the other side the link between latent states and words. The first connection represents the probability that a motion symbol generates the corresponding latent state while the second connection expresses the probability that a latent state generates a word. With this structure, the latent states layer symbolizes the association between motion symbols and words.

The motion language model and the natural language model are integrated by performing a computational search of sentences given a motion or for motions given a sentence. In this paper we focus only on the second direction. Furthermore, the usage of a bigram structure for the natural language model ensures a simple and efficient search appropriate for queries in both cases.

A more detailed explanation about both the motion language and the natural language model can be found in [17].

V. EXPERIMENTS

In order to test the potential applications of the proposed approach we acquired simple motion primitives labeled with adequate descriptions in a 2 DoF virtual scenario for a cooperative joint transportation task. We trained all different primitives using a single parametric dependency model, except for one primitive, used to test the applicability of the approach to unrelated motions. Using the descriptions and the learned PHMMs, we built both the corresponding natural language model and the motion language model. The presented results show the generalization properties of the system and its validity.

A. Experimental Setup

In a human-robot joint transportation task, due to the physical coupling, both partners must agree on the path to follow as they navigate through the environment. In case of disagreement, humans can easily communicate with each other using natural language in order to find a consensus and proceed with the task execution. However, finding an agreement implies the understanding of the description of motions through natural language. Such descriptions may also refer to environmental constraints. In this scenario, we designed a dataset in a simple environment in order to teach a robotic partner motion primitives with their respective description and environmental constraints that typically arise during this kind of interaction.

The training set used in the experiments is extracted with a haptic interface where the human applies forces in order to move the object along a desired trajectory, see Fig. V-A. The haptic interface consists of a two degrees-of-freedom linear-actuated device (*ThrustTube*) having a free-spinning handle at the grasping point. Attached to the handle, a force/torque sensor (*JR3*) measures the human force input that serves as

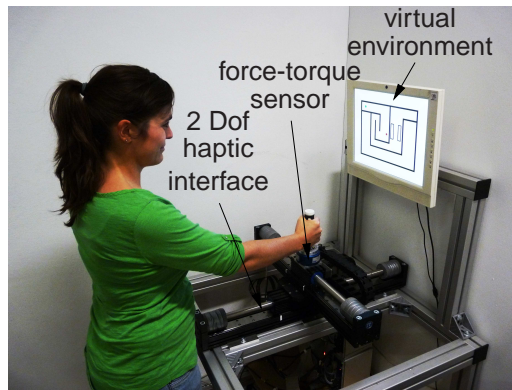


Fig. 4. Experimental setup.

the input for an admittance control scheme. A virtual scene is visually represented on a display placed on top of the interface showing a box in the middle.

Using this setup, we generated samples for 7 different primitives: "go straight", "turn right", "turn left", "go to the right of the box", "go to the left of the box", "go behind the box left" and "go behind the box right". For each primitive we produced 25 motions, combining the standard description with the adverbs "fast" and "slow" as well as changing the size of the box displayed in the scene with three different sizes, i.e. big= 3, middle= 2 and small= 1. Several exemplary motions from the dataset for "go behind the box right" and "turn left" are shown in Fig. 5. The motion data captured in task space was encoded in an invariant way, which, in the case of pure translation without orientation in 2 dimensions reduces to both the curvature and the velocity profile [14].

Given the description of the motions, the dictionary used for this scenario contains 13 different words. The parameters for the observations are therefore determined by the word parameters θ_l , a vector of size 13, and the environmental ones θ_e , given by the size of the box displayed in the scene. Considering the 7 different motion symbols and the sentences used for the descriptions, the natural language and the motion language model were built as explained in Section IV.

With these conditions, we trained all PHMMs sharing the same dependency function $f(\theta)$ as explained in Section III, except for the one representing "go behind the box left", which was trained apart. Each PHMM was trained with $N = 8$ states with $K = 1$ mixture components each. The GMM of generalized parameter function was trained with 8 mixture components. The following results show the expected motion for a query given to the system and using the generalized parameter model learnt for all primitives. The resulting motion is then reconstructed from the invariant representation back to the task coordinates.

B. Experimental Results

In order to show the general parametric model's generalization capabilities, the following queries given to the system are always referring to the primitive learned apart, "go behind the box left", which was not considered in $f(\theta)$.

The resulting motions for the queries "go behind the box

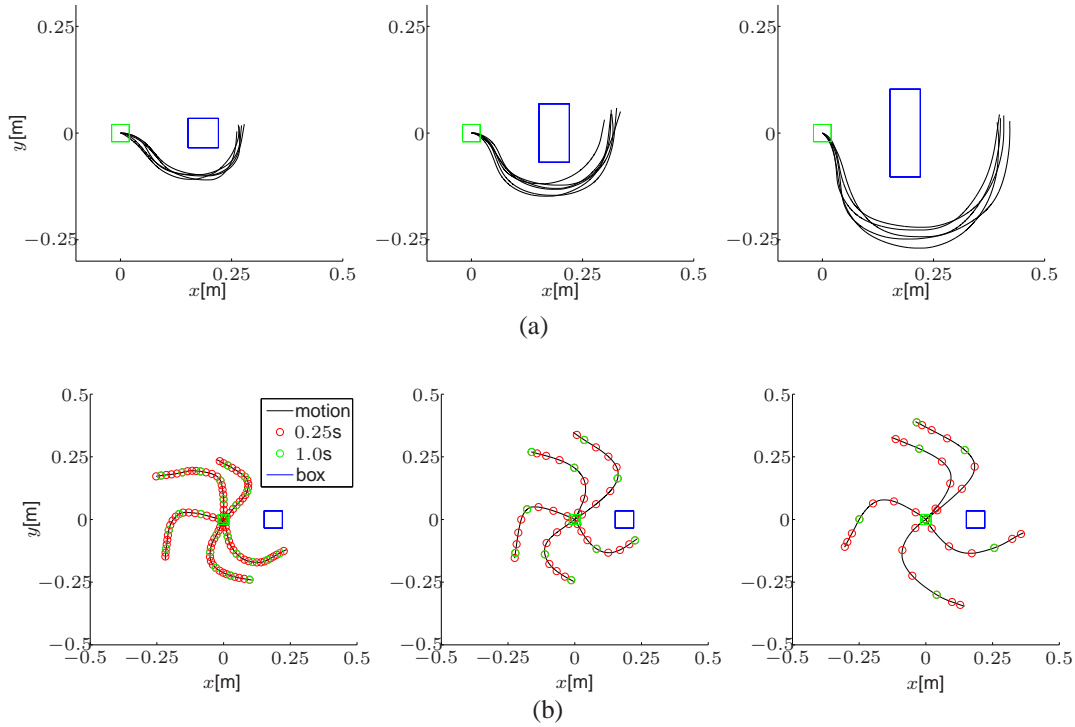


Fig. 5. Generated dataset for the primitives "go behind the box right" for box sizes 1, 2 and 3 on top. On the bottom, generated motions for "turn left" performed in three different speeds: slow, normal and fast.

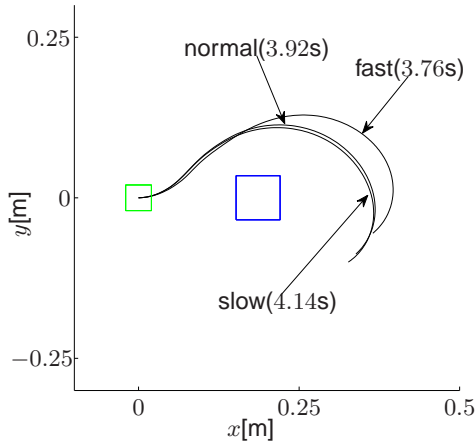


Fig. 6. Resulting motions for the queries "go behind the box left", "go behind the box left slow" and "go behind the box fast".

left", "go behind the box left slow" and "go behind the box fast" all considering a middle box size are shown in Fig. 6. The motion language model selected the motion symbol representing the primitive "go behind the box left" for all three cases. As shown by the execution times, the three resulting motions, while having very similar curvature profiles differ mainly on the execution time, even when the parameterization of "fast" or "slow" for this primitive was not included in the generalized parameter model.

For the queries "go behind the box left" with box sizes 0.8, 1, 2, 3 and 4, the resulting motions are shown in Fig. 7. As with the previous queries, the motion language model selected always the primitive "go behind the box left". In

this case, the queries consider parameter values which were even not in range of the training set and a motion that was not included in the generalized parameter model. However, the resulting motions show the intuitive response of the system which can even model the influence of the box in an abstract manner for a primitive not considered previously. While this result might be initially surprising, it is mainly due to the inclusion of the primitive "go to the left of the box" in the generalized parameter model which was trained for different box sizes. As both motion descriptions share many words ("go", "the", "box", "left"), the parameterization tends to extract most of the non-linear dependency from that model, whose motion profile is indeed similar to "go behind the box left", except for the end of the motion.

While the previous experiments show a proper response of the system to fully specified queries, the marginal influence of some parameters can also be extracted generating the expected mean posterior of their marginal distribution. In order to test the influence of the words like "fast" and "slow", we can calculate the parameterization of the model as the marginal distributions $P(\mathbf{f}(\theta)|\text{"fast"} = 1, \text{"slow"} = 0)$, $P(\mathbf{f}(\theta)|\text{"fast"} = 0, \text{"slow"} = 0)$ and $P(\mathbf{f}(\theta)|\text{"fast"} = 0, \text{"slow"} = 1)$. The results for these 3 queries applied to the primitive "go behind the box left" are shown in Fig. 8. As shown in the execution times, the essence of the adverbs is present as the fast motion is the fastest and the slow motion the slowest. However, it is also remarkable that in this case the motion profiles slightly differ showing an undesired response. This effect might be produced by the small amount of primitives considered in this scenario.

In summary, the proposed approach captures the influence

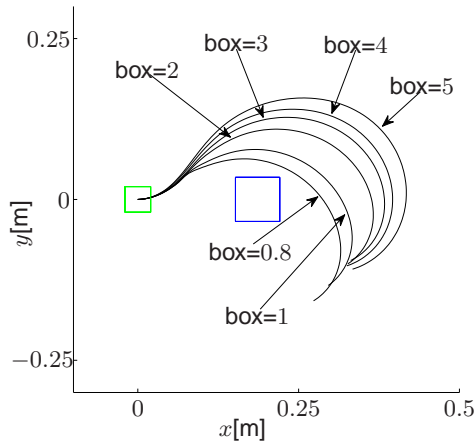


Fig. 7. Generated motions for the queries "go behind the box left" with box sizes 0.8, 1, 2, 3, 4 and 5.

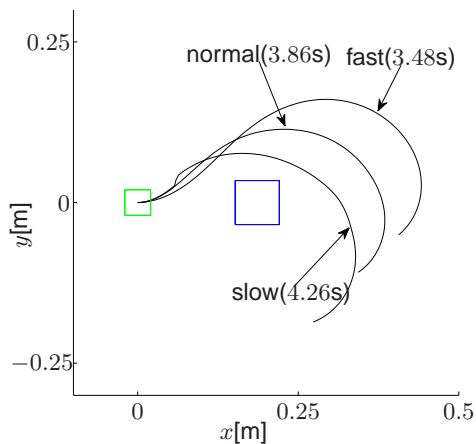


Fig. 8. Generated motions for the generated parameterization of the marginal probabilities of "slow" and "fast" given by $P(\mathbf{f}(\theta) | \text{"fast"} = 1, \text{"slow"} = 0)$, $P(\mathbf{f}(\theta) | \text{"fast"} = 0, \text{"slow"} = 0)$ and $P(\mathbf{f}(\theta) | \text{"fast"} = 0, \text{"slow"} = 1)$ applied to the primitive "go behind the box left".

of external parameters on motions using PHMMs. The motion language model extracts the most likely motion symbol given a natural language query and, as shown by the motion generation results, the parameterization of primitives not considered in the training using the generalized parameter model produces motions in agreement with their real meaning. Furthermore, the influence of isolated parameters is abstracted calculating their expected marginal distribution also producing similar results.

VI. CONCLUSION

In this paper we present a new approach to the problem of learning natural language together with motions and environmental properties. The proposed PHMM model captures the direct influence on motions of extrinsic elements such as adverbs from a description and environmental features. Furthermore, a generalized parameter model is obtained training together different motion primitives with the same parametric function. As shown by the presented results, the

proposed approach extracts the abstract meaning of external parameters and successfully applies its effect to motions not considered in the training. Combined with the natural language model and the motion language model, given a natural language query the presented system can also extract the most likely primitive from a motion symbol database and generate the expected motion.

The presented application shows the potential of the proposed approach for enhancing pHRI tasks through natural language communication. Testing the generalization capabilities for larger datasets including a more extensive dictionary and richer environmental features, as well as evaluating the system during disagreement in full scale scenarios during joint manipulation tasks is the matter of our future work.

ACKNOWLEDGMENTS

This research is partly supported by the DFG excellence initiative research cluster "Cognition for Technical Systems - CoTeSys".

REFERENCES

- [1] J. Medina, D. Lee, and S. Hirche, "An Incremental Learning and Prediction Framework and its Application in Cooperative Manipulation (submitted)," in *Proc. IEEE/ROS IROS*, 2011.
- [2] W. Takano, K. Yamane, T. Sugihara, K. Yamamoto, and Y. Nakamura, "Primitive communication based on motion recognition and generation with hierarchical mimesis model," in *ICRA*, 2006, pp. 3602–3609.
- [3] T. Ogata, S. Matsumoto, J. Tani, K. Komatani, and H. G. Okuno, "Human-robot cooperation using quasi-symbols generated by rnnpb model," in *ICRA*, 2007, pp. 2156–2161.
- [4] T. Miyashita, T. Tajika, H. Ishiguro, K. Kogure, and N. Hagita, "Haptic communication between humans and robots," in *ISRR*, 2005, pp. 525–536.
- [5] Y. Sugita and J. Tani, "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes," *ADAPTIVE BEHAVIOR*, vol. 13, pp. 33–52, 2005.
- [6] K. Sugiura, N. Iwahashi, H. Kashioka, and S. Nakamura, "Active learning of confidence measure function in robot language acquisition framework," in *IROS*, 2010, pp. 1774–1779.
- [7] W. Takano, K. Yamane, and Y. Nakamura, "Capture database through symbolization, recognition and generation of motion patterns," in *ICRA*, 2007, pp. 3092–3097.
- [8] W. Takano and Y. Nakamura, "Statistically integrated semiotics that enables mutual inference between linguistic and behavioral symbols for humanoid robots," in *ICRA*, 2009, pp. 646–652.
- [9] —, "Associative processes between behavioral symbols and a large scale language model," in *ICRA*, 2010, pp. 2404–2409.
- [10] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot Programming by Demonstration," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1371–1394.
- [11] D. Lee and C. Ott, "Incremental kinesthetic teaching of motion primitives using the motion refinement tube," *Auton. Robot.*, pp. 1–17, 2011.
- [12] A. Wilson and A. Bobick, "Parametric hidden markov models for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 884–900, 1999.
- [13] D. Herzog and V. Krüger, "Statistical and geometrical approaches to visual motion analysis," ch. Recognition and Synthesis of Human Movements by Parametric HMMs.
- [14] J. D. Schutter, "Invariant Description of Rigid Body Motion Trajectories," *ASME Journal of Mechanisms and Robotics*, vol. 2, pp. 011 004–1–011 004–9, 2010.
- [15] J. Bilmes, "A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," U.C. Berkeley, Tech. Rep., 1997.
- [16] D. Cohn, Z. Ghahramani, and M. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.
- [17] W. Takano and Y. Nakamura, "Bigram-based natural language model and statistical motion symbol model for scalable language of humanoid robots," in *ICRA*, 2012.