The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Generating Robot Gesture Using a Virtual Agent Framework

Maha Salem, Stefan Kopp, Ipke Wachsmuth, Frank Joublin

Abstract— **One of the crucial aspects in building sociable, communicative robots is to endow them with expressive non-verbal behaviors. Gesture is one such behavior, frequently used by human speakers to illustrate what they express in speech. The production of gestures, however, poses a number of challenges with regard to motor control for arbitrary, expressive hand-arm movement and its coordination with other interaction modalities. We describe an approach to enable the humanoid robot ASIMO to flexibly produce communicative gestures at run-time, building upon the Articulated Communicator Engine (ACE) that was developed to allow virtual agents to realize planned behavior representations on the spot. We present a control architecture that tightly couples ACE with ASIMO's perceptuo-motor system for multi-modal scheduling. In this way, we combine conceptual representation and planning with motor control primitives for meaningful arm movements of a physical robot body. First results of realized gesture representations are presented and discussed.**

I. INTRODUCTION

Lifelike acting in a social robot evokes social communicative attributions to the robot and thereby conveys intentionality. That is, the robot makes the human interaction partner believe that it has, e.g., internal states, communicative intent, beliefs and desires [4]. To induce such beliefs, a robot companion should produce social cues. Forming an integral part of human communication, hand and arm gestures are primary candidates for extending the communicative capabilities of social robots. Non-verbal expression via gesture is frequently used by human speakers to emphasize, supplement or even complement what they express in speech. Pointing to objects or giving spatial direction are good examples of how information can be conveyed in this manner. This additional expressiveness is an important feature of social interaction to which humans are known to be well attentive. Similarly, humanoid robots that are intended to engage in natural and fluent human-robot interaction should produce communicative gestures for comprehensible and believable behavior.

In contrast to task-oriented movements like reaching or grasping, human gestures are derived to a certain extent from some kind of internal representation of 'shape' [11], especially when iconic or metaphoric gestures are used. Such characteristic shape and dynamical properties exhibited by gestural movement allow humans to distinguish them from subsidiary movements and to recognize them as meaningful non-verbal behavior [24]. As a consequence, the generation of gestures for artificial humanoid bodies demands a high degree of control and flexibility concerning shape and time properties of the gesture, while ensuring a natural appearance of the movement. Ideally, if such non-verbal behaviors are to be realized, they have to be derived from conceptual, to-be-communicated information.

The present paper focuses on the implementation of communicative gestures which have to meet the aforementioned constraints. The overall objective of this research is to enable a physical robot to flexibly produce speech and co-verbal gesture at run-time and to subsequently evaluate the resulting communicative behavior in human-robot interaction studies. For this, we explore how we can transfer existing concepts from the domain of virtual conversational agents to the platform of a humanoid robot. In [21], we address the production of speech as a further output modality and its synchronization with gesture. A future aspect of this work will incorporate an evaluation of the generated multi-modal robot behavior.

II. RELATED WORK

Up to now, the generation together with the evaluation of the effects of robot gesture is largely unexplored. In traditional robotics, recognition rather than synthesis of gesture has mainly been brought into focus. In the few existing cases of gesture synthesis, however, models typically denote object manipulation fulfilling little or no communicative function, e.g. [2]. Furthermore, gesture generation is often based on the recognition of previously perceived gestures, thereby focusing on imitation learning, e.g. [1]. In most cases in which robot gesture is actually generated with a communicative intent, these arm movements are not produced at run-time, but are pre-recorded for demonstration purposes, e.g. [23] and [7].

Crucially, many approaches are realized on less sophisticated platforms with less complex robot bodies (e.g., less degrees of freedom (DOF), limited mobility, etc.) that show no or only few humanoid traits. However, it is not only the behavior but also the appearance of a robot that affects the way human-robot interaction is experienced [19]. Consequently, the importance of the robot's design should not be underestimated if the intention is to ultimately use it as a research platform, e.g., to study the effect of robot gesture on humans. MacDorman and Ishiguro consider android robots a key testing ground for social, cognitive, and neuroscientific theories, providing an experimental apparatus that can be controlled more precisely than any human actor [16].

M. Salem is at the Research Institute for Cognition and Robotics, Bielefeld, Germany msalem@cor-lab.uni-bielefeld.de
S. Kopp is at the Sociable Agents Group, Bielefeld University, Germany skopp@techfak.uni-bielefeld.de
I. Wachsmuth is at the Artificial Intelligence Group, Bielefeld University, Germany ipke@techfak.uni-bielefeld.de
F. Joublin is at the Honda Research Institute Europe, Offenbach, Germany frank.joublin@honda-ri.de

This is in line with initial results which indicate that only robots strongly resembling humans, e.g., being equipped with a head, two arms and two legs, can elicit the broad spectrum of responses that people typically direct toward each other. These findings highlight the benefits of using the humanoid Honda robot ASIMO as a research platform for the evaluation of human-robot interaction (HRI). Although the present work mainly focuses on the technical aspects of robot gesture generation, a major objective of this research is the assessment of such non-verbal robot behavior in various HRI scenarios.

While the generation of gesture, especially together with synchronized speech, is relatively novel to the research area of robotics, it has already been addressed in various ways within the domain of virtual humanoid agents. Over a decade ago, Cassell et al. introduced the *REA* system [3] employing a conversational humanoid agent that plays the role of a real estate salesperson. Gibet et al. [5] generate and animate sign-language from script-like specifications, resulting in fairly natural movement characteristics. Even in this domain, however, most existing systems neglect the meaning a gesture conveys or simplify matters using lexicons of words and canned non-verbal behaviors in the form of pre-produced gestures [8]. On the contrary, the framework underlying the virtual agent Max [12] aims at an integrated architecture in which the planning of both content and form across both modalities is coupled [10], hence taking into account the meaning conveyed in non-verbal utterances. Generally speaking, computational approaches to generating multi-modal behavior can be modeled in terms of three consecutive tasks [20]: firstly, determining *what* to convey (i.e., content planning); secondly, determining *how* to convey it (i.e., micro-planning); finally, realizing the planned behaviors (i.e., surface realization). Even though the Articulated Communicator Engine (ACE) itself operates at the end of the generation pipeline, i.e., on the surface realization layer, the overall system used for Max also provides an integrated content planning and micro-planning framework [10]. In the present paper, however, we focus on ACE, as it marks the starting point required for the interface endowing the robot ASIMO with similar communicative behavior.

III. UNDERLYING MULTI-MODAL PRODUCTION MODEL

The presented approach is predicated on descriptions specifying the outer form of the multi-modal utterances that are to be communicated. For this purpose, the XML-based Multi-modal Utterance Representation Markup Language (MURML [13]) is used to specify verbal utterances in combination with co-verbal gestures. These, in turn, are either explicitly described in terms of form features (i.e., the posture designated for the gesture stroke) using feature-based MURML specifications or, alternatively, can be defined as keyframe animations. The latter are based on the specification of different 'key postures' for each keyframe, which describe the current state of each joint as part of the overall gesture movement pattern and are interpolated at runtime. In ACE, keyframe animations can be either defined manually or derived from motion capturing of a human demonstrator, allowing the real-time animation of virtual agents. In both feature-based and keyframe-based MURML descriptions gesture affiliation to dedicated linguistic elements is specified based on matching time identifiers. Fig. 1 illustrates a sample feature-based MURML specification that can be used as input for speech-gesture production in ACE and the resulting gesture executed by the virtual agent Max. For more information on MURML see [13].

The concept underlying the multi-modal production model acts on an empirically suggested assumption referred to as the *segmentation hypothesis* [18]. It claims that the production of continuous speech and gesture is organized in successive segments. Each of these segments represents a single idea unit which is referred to as a *chunk* of speech-gesture production. A chunk, in turn, consists of an intonation phrase and a co-expressive gesture phrase, concertedly conveying a prominent concept [12]. Levelt [15] defines intonation phrases to represent units over which the phonological structure of continuous speech is organized. Accordingly, Kendon [9] describes gesture phrases as units of gestural movement comprising one or more subsequent phases: preparation, stroke, retraction, hold. Although MURML allows for the specification of both intonation and gesture phrases to produce multi-modal output in ACE, the present paper focuses on the generation of gestures only.

Gesture motor control is realized hierarchically in ACE: During higher-level planning, the motor planner is provided with timed form features as annotated in the MURML specification. This information is then passed on to independent motor control modules. The idea behind this functional-anatomical decomposition of motor control is to break down the complex control problem into solvable sub-problems. ACE [12] provides specific motor planning modules for the arms, the wrists, and the hands which, in turn, instantiate local motor programs (LMPs). These are used to animate required sub-movements. LMPs operate within a limited set of DOFs and over a designated period of time. For the motion of each limb, an abstract motor control program (MCP) coordinates and synchronizes the concurrently running LMPs, gearing towards an overall solution to the control problem.

```
<definition><utterance>
 <specification>
 And now take the object <time id="t1" chunkborder="true"/>
 and make it <time id="t2"/> this big. <time id="t3"/>
 </specification>
 <behaviorspec>
  <gesture id="gesture_1" scope="hand">
   <affiliate onset="t2" end="t3" focus="this"/>
    <constraints>
     <symmetrical dominant="right_arm" symmetry="SymMS">
      <parallel>
       <static slot="HandShape" value="BSflat(FBround all o)"/>
       <static slot="ExtFingerOrientation" value="DirA"/>
       <static slot="PalmOrientation" value="DirL"/>
       <static slot="HandLocation" value="LocChest LocCenterRight LocNorm"/>
      </parallel>
     </symmetrical>
    </constraints>
  </gesture>
 </behaviorspec>
</utterance></definition>
```

Fig. 1. Example of a feature-based MURML specification for multi-modal utterances in ACE and the resulting gesture.

The top-level control of the ACE framework, however, does not attend to how such sub-movements are controlled. To ensure an effective interplay of the LMPs involved in a MCP, the planning modules arrange them in a controller network which defines their potential interdependencies for mutual (de-)activation. LMPs are able to transfer activation between themselves and their predecessors or successors to allow for context-dependent gesture transitions. Thus they can activate or deactivate themselves at run-time depending on feedback information on current movement conditions.

On-line timing of gestures is achieved on-the-fly by the ACE engine as follows: The ACE scheduler retrieves timing information about the synthetic speech at the millisecond level and defines the start and the end of the gesture stroke accordingly. These temporal constraints (e.g., how long the hand has to form a certain shape) are propagated down to each single gesture component automatically. Subsequently, the motor planner creates the LMPs that meet both the temporal and the form constraints. The second aspect of scheduling, i.e., the decision to skip preparation or retraction phases, automatically results from the interplay of motor programs at run-time. Motor programs monitor the body's current movements and activate themselves to realize the planned gesture stroke as scheduled. A retraction phase is skipped whenever the motor program of the following gesture takes over the control of the effectors from the preceding program. This on-line scheduling leads to fluent and continuous movements.

IV. Control Architecture for Robot Gesture

Our research aim is to endow the robot ASIMO with 'conceptual motorics'. In order to achieve this, a robot control architecture that combines conceptual representation and planning with motor control primitives for arm and hand movements is required. This poses several challenges: since ACE was originally designed for a virtual agent application, it fails to adequately account for certain physical restrictions such as motor states, maximum velocity, strict collision avoidance, variation in DOFs, etc. Ultimately, we need to tackle these challenges when transferring our virtual agent framework ACE to the physical robot ASIMO.

Models of human motor control commonly exhibit a hierarchical structure [22], representing global aspects of a movement as an abstract goal at the highest level. Control is passed down through lower levels until all choices about which motor units to use are made. Latash [14] suggests that planning of movements is directly performed in terms of kinematics in the external task space rather than in the more complex joint space. His proposed general scheme of motor control forms the basis of our architecture for robot motor control; it incorporates three different levels: Firstly, the planning of a movement in which the abstract goal of the intended movement is described in terms of an internally simulated trajectory. Secondly, this trajectory is then translated into motor variables and commands which control the lower structures, resulting in virtual trajectories. These, in turn, partially encode certain properties of the movement, e.g., specific patterns of transition from initial to final position. Finally, the third level serves to execute these commands at the lowest level, resulting in a movement that ideally matches the simulated trajectory.

Control Parameters: Our approach to robot control in combination with ACE allows for using two different kinds of control parameters to drive ASIMO: joint angles or task-space coordinates and orientation. The first method involves an extraction of the joint angles already from the ACE kinematic body model, which are then mapped onto the robot body model. The second method amounts to using ACE to formulate a trajectory in terms of effector targets and orientations in task space, based on which a joint space description using the standard whole body motion (WBM) controller [6] for ASIMO can be derived. WBM allows to control all DOFs of the humanoid robot based on given end-effector targets, thus providing a flexible method to control upper body movement by only specifying relevant task dimensions selectively in real-time. In doing so, task-specific command elements can be assigned to the command vector at any given instant. This allows the system to control one or multiple effectors while generating a smooth and natural movement. Redundancies are optimized with regard to joint limit avoidance and self-collision avoidance.

We opted for the second approach, task-space control, for several reasons. First, since ACE was originally designed for a virtual agent application, it does not entirely account for certain physical restrictions such as collision avoidance, which may lead to joint states that are not feasible on the robot. In contrast, solving IK using ASIMO's internally implemented WBM controller ensures safer postures for the robot. Further, task-space control is in line with results from human perception studies suggesting that humans largely track the hand or end-points of one another's movement, even if the movement is performed with the entire arm [17]. Evidently, even with a deviation of joint angles the form and meaning of a gesture can still be conveyed.

However, the given constraints as well as velocity limits and motor states can affect the performance of the robot. Thus, the inner states represented within the kinematic body model in ACE may deviate from the actual motor states of the physical robot during run-time. For this reason, a bi-directional interface using both efferent actuator control signals and afferent sensory feedback is required. This can be realized by a feedback loop that updates the internal model of ASIMO in WBM as well as the kinematic body model coupled to ACE at a sample rate r. For successful integration, this process needs to synchronize two competing sample rates: that of the ACE engine, and that of the WBM software controlling ASIMO. Fig. 2 illustrates the control architecture we employ for robot gesture based on the ACE framework.

Sampling Rate: Another crucial issue when connecting the ACE framework to the ASIMO control layer is the rate at which control parameters are passed from the former to the latter (sampling rate). A number of alternative mapping rates might be employed:

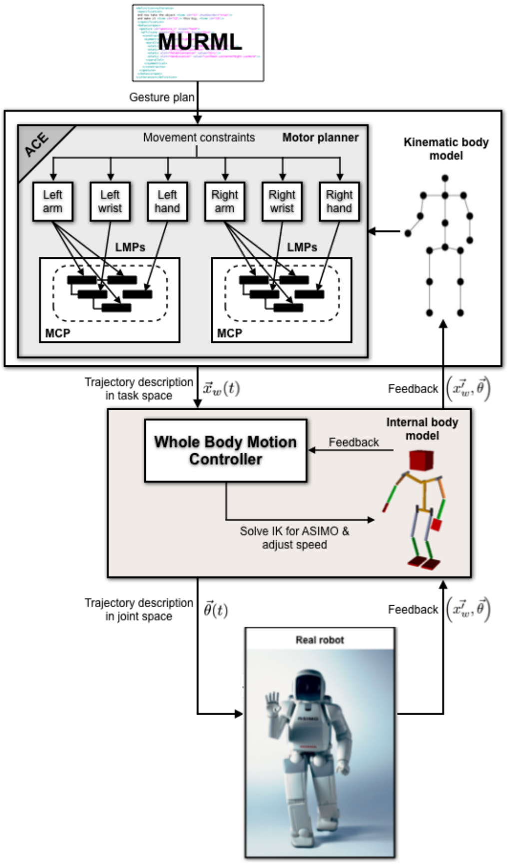


Fig. 2. Control architecture for the realization of robot gesture.

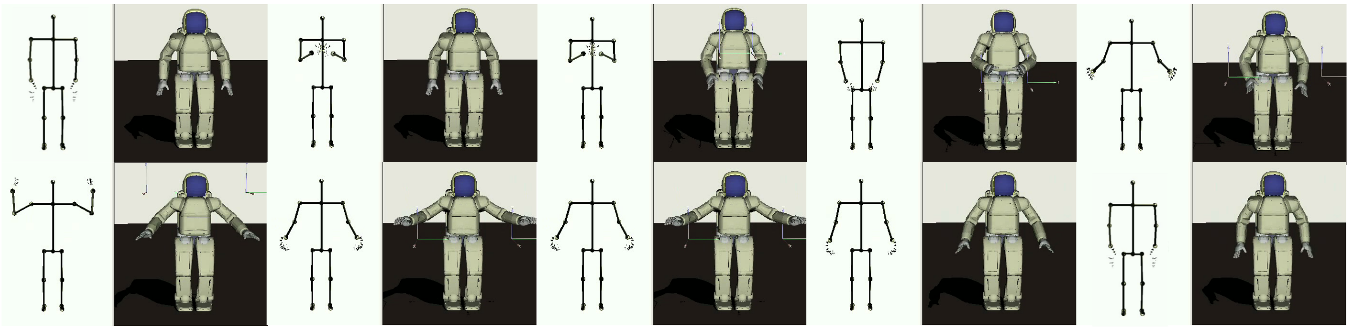
1) sampling at each n-th frame: ACE would send control parameters at a fixed rate to ASIMO's WBM controller;
2) sampling only at target positions: ACE would send only the end positions or orientations of movement segments and delegates the robot movement generation entirely to ASIMO's WBM controller;
3) adaptive sampling rate: ACE would "tether" WBM using different sampling rates, ranging from one sample per frame to taking only the end positions.

In our current setup, we employ the first method with a maximal sampling rate, i.e., each successive frame of the movement trajectory is sampled and transmitted to the robot controller (n=1). Given a frame rate of 20 frames per second (flexibly adjustable with ACE), this can result in a large number of sample points which, in turn, ensures that ASIMO closely follows the possibly complex trajectory planned by ACE. In the following section, we will present results obtained with this method. Alternatively, the third strategy would allow for adjusting the sampling rate depending on the trajectory's complexity, which may well vary from simple straight movements (e.g., for gesture preparation) to complex curved shapes for the gesture stroke phase. If the trajectory is linear, then we can expect that strategy 2 above might serve as the best mechanism since only distance information would likely be required. If, on the other hand, the trajectory is complex, we can expect that strategy 1 would be optimal, since a sequence of small movement vectors would likely be required to guide the robot controller. In a realistic scenario, in which a particular gesture is formed from several different types of sub-movement, a composition of both strategies as possible with an adaptive sampling rate would become optimal. This is a point of future investigation.

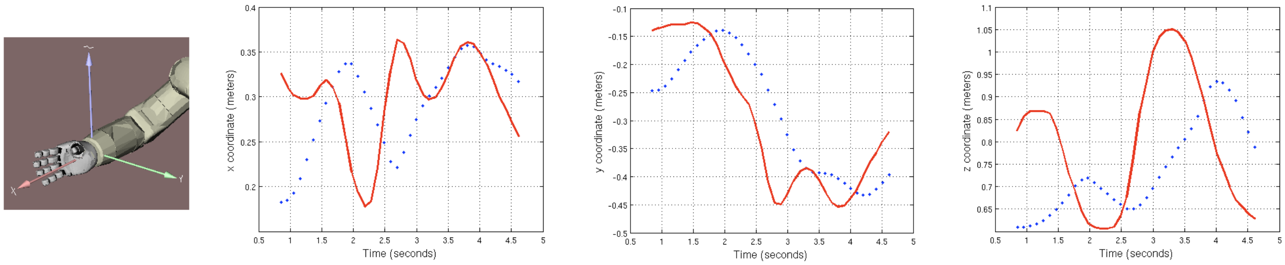
V. First Results and Discussion

Our first results were produced in a feed-forward manner constantly transmitting commands of the wrist position from ACE to ASIMO at a sample rate of 20 frames per second. Inverse Kinematics is then solved using the whole body motion controller provided by ASIMO. Fig. 3(a) illustrates a sample gesture based on a MURML keyframe animation realized in our current framework, Fig. 4(a) illustrates the execution of a feature-based MURML specification. We use the ASIMO simulation software which is restricted by the original physical constraints and visualizes fairly accurately the expected movement behavior of the real ASIMO robot. The kinematic body model representing the internal state of ACE at each time step is displayed next to the robot. The screen-shots reveal how ASIMO can only perform the movements with a remarkable delay caused by more restrictive velocity limits. Figs. 3(b) and 4(b)I. plot each dimension of the wrist position of both the ACE body model and ASIMO against time to illustrate this observation. Finally, Fig. 4 (b) II. plots the Euclidean distance between the current wrist position of the ASIMO robot and the target position marked by the ACE body model at each time step during execution of the given gesture. It indicates that the greatest discrepancy occurs in the preparation and the retraction phase of a gesture in which abrupt shifts in direction take place.

Despite the general limitation in speed, these findings substantiate the feasibility of the proposed approach. Generally, arbitrary MURML-based gesture representations – both feature-based descriptions as well as keyframe animations which can be optionally derived from human motion capturing data – can now be realized using the current framework. Extensive tests with multiple various gesture representations (including both one-armed and two-armed movements) performed on the physical ASIMO robot further revealed that task-space control (i.e., disregarding the joint angles as generated in ACE) does not impair the overall shape of a gesture. Consequently, controlling the robot via task space commands turns out to be an appropriate and safe way to generate arm movement behavior. However, if this framework is to be used for producing multi-modal utterances incorporating synchronized speech, the timing of gesture generation will be important. We will need to find ways to overcome the difference in time required by the internal kinematic body model in ACE on the one hand and by ASIMO's physically constrained body on the other. Our approach is to realize this by extending the cross-modal adaptation mechanisms applied in ACE to allow for a finer mutual adaptation between robot gesture and speech. This requires the incorporation of a forward model to predict the

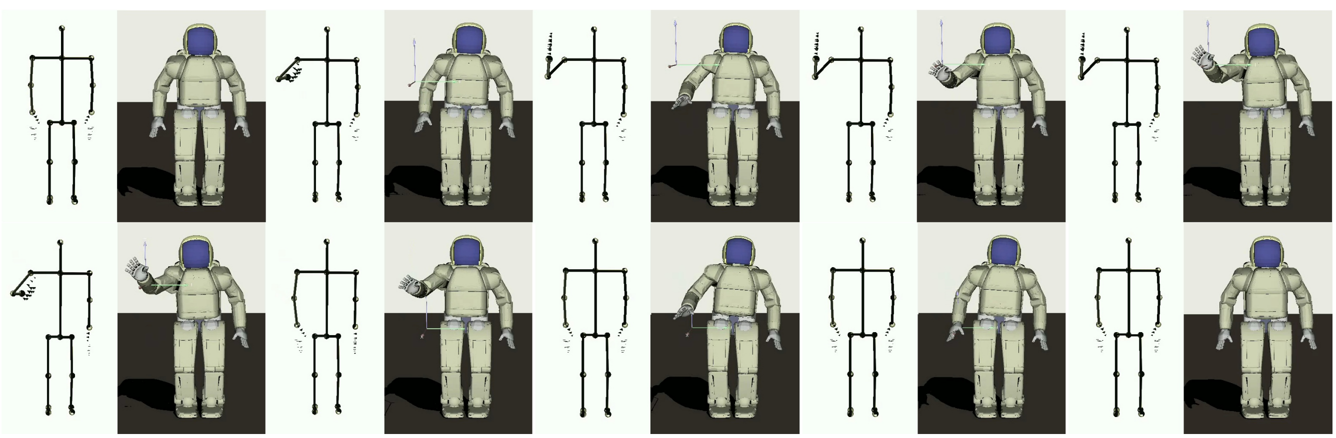


(a) Comparison between the ACE kinematic body model and the ASIMO robot (left to right, top-down, 0.6 sec intervals).

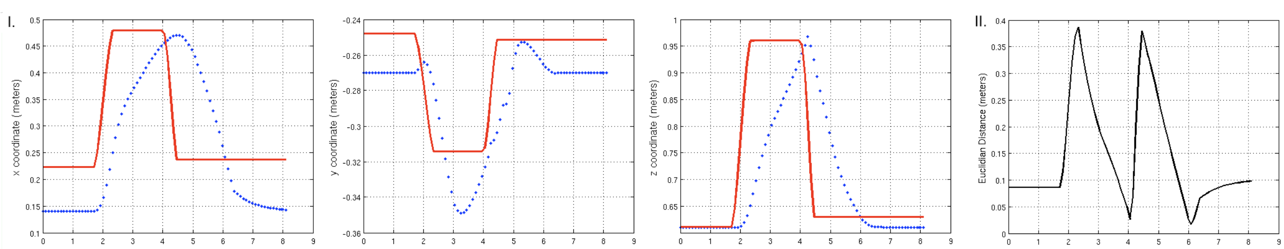


(b) Plot of the x-, y-, and z-coordinate of the right wrist positions of the ACE body model (solid) and ASIMO (dotted) during gesture execution.

Fig. 3. Sample gesture using a MURML keyframe animation as realized in the current framework.



(a) Comparison between the ACE kinematic body model and the ASIMO robot (left to right, top-down, 0.4 sec intervals).



(b) I. Plot of the x-, y-, and z-coordinate of the right wrist positions of the ACE body model (solid) and ASIMO (dotted) during gesture execution; II. Euclidean distance between current wrist position of ASIMO robot and target position marked by ACE body model over time.

Fig. 4. Sample gesture using a feature-based MURML specification as realized in the current framework.

time needed for gesture preparation. Additionally, predicted values must be controlled and, if necessary, adjusted based on constantly updated feedback information on the robot state.

VI. Conclusion and Future Work

We presented a robot control architecture to endow the humanoid robot ASIMO with flexibly produced gestures at run-time. The proposed framework is based on a speech and gesture production model originally developed for a virtual human agent. The Articulated Communicator Engine (ACE) is one of the most sophisticated multi-modal schedulers and allows for an on-the-spot production of flexibly planned behavior representations. Our framework combines conceptual, XML-based representation and planning with motor control primitives for arm and hand movements. This way, predefined canned behaviors can be replaced by conceptual motorics generated at run-time. Re-employing ACE as an underlying action generation engine enables us to draw upon a tight coupling of ASIMO's perceptuo-motor system with multi-modal scheduling. While our first results were produced in a feed-forward manner, the proposed robot control architecture will eventually make use of both efferent control signals and afferent feedback.

The need to ensure temporal and semantic coherence of communicative behavior by meeting strict synchrony constraints imposed by a further output modality, namely speech, will present a main challenge to our framework in the future. Clearly, the generation of finely synchronized multi-modal utterances proves to be more demanding when realized on a robot with a physically constrained body than for an animated virtual agent, especially when communicative signals must be produced at run-time. Currently, the ACE engine achieves synchrony mainly by gesture adaptation to structure and timing of speech, obtaining absolute time information at phoneme level. However, to tackle this new dimension of requirements, the cross-modal adaptation mechanisms applied in ACE will have to be extended to allow for a finer mutual adaptation between robot gesture and speech.

As yet, the generation together with the evaluation of the effects of robot gesture is largely unexplored. Our results are a step towards shedding light on conceptual motorics in robotic agents. Crucially, they provide a proof of concept, highlighting the feasibility of the approach while also demonstrating the direction for our future research. Once our robot control architecture has been extended to fully account for both speech and co-verbal gesture as well as the fine synchronization of the two modalities, the framework must be evaluated against different performance evaluation criteria. Furthermore, it will be assessed in a human-robot interaction scenario, in which both a human interaction partner and robot perform a joint task. In a suitable scenario a human subject will be asked to identify and move differently shaped and sized objects which will be referred to by ASIMO using a variety of different gestures. In conclusion, the implementation and evaluation of our robot control architecture realized on ASIMO will provide new insights into human perception and understanding of gestural machine behaviors and how to use these in designing more human-like robot communication.

VII. ACKNOWLEDGMENTS

The research project "Conceptual Motorics" is supported by the Honda Research Institute Europe.

REFERENCES

[1] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Robot Programming by Demonstration. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*, pages 1371–1394. Springer, Secaucus, NJ, USA, 2008.

[2] S. Calinon and A. Billard. Learning of Gestures by Imitation in a Humanoid Robot. In K. Dautenhahn and C.L. Nehaniv, editors, *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, pages 153–177. Cambridge University Press, 2007.

[3] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjálmsson, and H. Yan. Human Conversation as a System Framework: Desigining Embodied Conversational Agents. In *Embodied Conversational Agents*, pages 29–63. MIT Press: Cambridge, MA, 2000.

[4] K. Dautenhahn. Ants Don't Have Friends - Thoughts on Socially Intelligent Agents. In *AAAI Technical Report FS-97-02*, 1997.

[5] S. Gibet, T. Lebourque, and P.-F. Marteau. High-Level Specification and Animation of Communicative Gestures. *Journal of Visual Languages and Computing*, 12(6):657–687, 2001.

[6] M. Gienger, H. Janßen, and S. Goerick. Task-Oriented Whole Body Motion for Humanoid Robots. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan, 2005.

[7] J.F. Gorostiza, R. Barber, A.M. Khamis, M. Malfaz, R. Pacheco, R. Rivas, A. Corrales, E. Delgado, and M.A. Salichs. Multimodal Human-Robot Interaction Framework for a Personal Robot. In *RO-MAN 06: Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006.

[8] B. Hartmann, M. Mancini, and C. Pelachaud. Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. *Gesture in Human-Computer Interaction and Simulation*, 2005.

[9] A. Kendon. Gesticulation and Speech: Two Aspects of the Process of Utterance. *The Relationship of Verbal and Non-verbal Communication*, pages 207–227, 1980.

[10] S. Kopp, K. Bergmann, and I. Wachsmuth. Multimodal Communication from Multimodal Thinking - Towards an Integrated Model of Speech and Gesture Production. *Semantic Computing*, 2(1):115–136, 2008.

[11] S. Kopp and I. Wachsmuth. A Knowledge-based Approach for Lifelike Gesture Animation. In W. Horn, editor, *ECAI 2000 - Proceedings of the 14th European Conference on Artificial Intelligence*, pages 663–667, Amsterdam, 2000. IOS Press.

[12] S. Kopp and I. Wachsmuth. Synthesizing Multimodal Utterances for Conversational Agents. *Computer Animation and Virtual Worlds*, 15(1):39–52, 2004.

[13] A. Kranstedt, S. Kopp, and I. Wachsmuth. MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents. In *Proceedings of the AAMAS02 Workshop on Embodied Conversational Agents - let's specify and evaluate them*, Bologna, Italy, July 2002.

[14] M.L. Latash. *Control of Human Movement*. Human Kinetics Publishers, Champaign, IL, USA, 1993.

[15] W.J. Levelt. *Speaking*. MIT Press, Cambridge, MA, 1989.

[16] K.F. Macdorman and H. Ishiguro. The Uncanny Advantage of Using Androids in Cognitive and Social Science Research. *Interaction Studies*, 7(3):297–337, 2006.

[17] Maja J. Mataric and Marc Pomplun. Fixation Behavior in Observation and Imitation of Human Movement. *Cognitive Brain Research*, 7(2):191 – 202, 1998.

[18] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.

[19] T. Minato, M. Shimada, H. Ishiguro, and S. Itakura. Development of an Android Robot for Studying Human-Robot Interaction. *Innovations in Applied Artificial Intelligence*, pages 424–434, 2004.

[20] E. Reiter and R. Dale. *Building Natural Language Generation Systems*. Cambridge Univ. Press, 2000.

[21] M. Salem, S. Kopp, I. Wachsmuth, and F. Joublin. Towards an Integrated Model of Speech and Gesture Production for Multi-Modal Robot Behavior. In *Proceedings of the 2010 IEEE International Symposium on Robot and Human Interactive Communication*, 2010.

[22] M.R. Sheridan. Planning and Controlling Simple Movements. *The Psychology of Human Movement*, pages 47–82, 1984.

[23] C.L. Sidner, C. Lee, and N. Lesh. The Role of Dialog in Human Robot Interaction. In *International Workshop on Language Understanding and Agents for Real World Interaction*, 2003.

[24] I. Wachsmuth and S. Kopp. Lifelike Gesture Synthesis and Timing for Conversational Agents. In I. Wachsmuth and T. Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, LNAI 2298, pages 120–133, Berlin, 2002. Springer.