

Model-based Animation of Coverbal Gesture

Stefan Kopp & Ipke Wachsmuth
Artificial Intelligence Group
Faculty of Technology, University of Bielefeld
D-33594 Bielefeld, Germany
e-mail: {skopp, ipke}@techfak.uni-bielefeld.de

Abstract

Virtual conversational agents are supposed to combine speech with nonverbal modalities for intelligible and believable utterances. However, the automatic synthesis of coverbal gestures still struggles with several problems like naturalness in procedurally generated animations, flexibility in pre-defined movements, and synchronization with speech. In this paper, we focus on generating complex multimodal utterances including gesture and speech from XML-based descriptions of their overt form. We describe a coordination model that reproduces co-articulation and transition effects in both modalities. In particular, an efficient kinematic approach to creating gesture animations from shape specifications is presented, which provides fine adaptation to temporal constraints that are imposed by cross-modal synchrony.

1. Introduction

Techniques from artificial intelligence, computer animation, and human-computer-interaction are recently converging in the growing field of embodied conversational agents [4]. Such agents are envisioned to have the same properties as humans in face-to-face conversation, including the ability to generate simultaneous verbal and nonverbal behaviors. This includes *coverbal* gestures that, in humans, are produced automatically and unconsciously during speech as part of an integrated utterance. Both modalities are tightly coupled such that synchronies w.r.t. semantics, pragmatics, and even timing can be observed [13], i.e., gestures are influenced by the communicative intent and the verbal utterance in many ways. In consequence, the animation of coverbal gestures requires a high degree of control and flexibility w.r.t. a gesture's shape and time properties while at the same time ensuring naturalness of movement. The demand for realism and real-time capability in multimodal agents has most often led to gestural behaviors which are either cap-

tured from real humans or manually predefined to a large extent. Yet, the employed techniques suffer from limited flexibility when it comes to adjusting a gesture's timing to accompanying speech (cf. [2, 6]) or concatenating them to continuous motion.

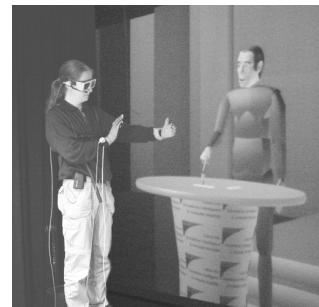


Figure 1. Multimodal interaction with Max.

In our lab, the anthropomorphic agent *Max* (see Fig. 1 for the overall scenario) acts as an assembly expert in an immersive 3D virtual environment for simulated construction and design tasks. The agent demonstrates assembly procedures to the user by combining facial and upper limb gestures with spoken utterances. Our work on synthesizing *Max*'s utterances, however, focusses on creating synchronized gestural and verbal behaviors solely from application-independent descriptions of their outer form. The gesture animation process builds on a hierarchical model of planning and controlling the upper-limb movements of an articulated figure, the planning stages being explained in [10]. After discussing related work in the next section, Section 3 describes how to specify the utterances of our agent. Section 4 explains our approach to motion control in creating appropriate gesture animations in real-time. The overall coordination framework for producing fluent, complex multimodal utterances with multiple verbal and gestural parts is explained in Section 5.

2. Related work

Most work concerning the animation of lifelike gesture in embodied agents relies on concatenating predefined motion elements which are drawn from static libraries. In conversational agents, communicative acts are usually mapped during motion planning onto behaviors identified with mostly stereotyped movements [17, 2]. Such motion primitives can be slightly parameterized and combined to form more complex movements, e.g., by means of high-level script languages [15]. In the Animated Conversation [3] and REA system [2] as well as in the recent BEAT system [5], Cassel et al. succeeded in predicting the timing of gesture animations such that the stroke coincides with the nuclear stress in speech. However, Cassell [1] states that the problem of creating final gesture animations and synchronizing them with speech has not been solved so far, “due in part to the difficulty of reconciling the demands of graphics and speech synthesis software” (p. 16). To some extent, this can be ascribed to the lack of sufficient means of modulating, e.g., shortening and stretching, single gesture phases (cf. [5]) while preserving human movement characteristics.

A fully automatic creation of upper limb movements by means of applying appropriate control models was targeted by only few researchers. Approaches based on control algorithms in dynamic simulations or optimization criteria provide a high level of control and may lead to physically realistic movements. However, these techniques suffer from difficulties in formulating appropriate control schemes for highly articulated figures and immense computational cost. Koga et al. [9] proposed a purely kinematic model for simulating pre-planned arm movements for grasping and manipulating objects. In particular, this work succeeded in applying findings from neurophysiology to create natural arm postures. Gibet et al. [7] apply generic error-correcting controllers for generating sign language from script-like specifications. These models succeeded in simulating natural movement characteristics to some extent but did not focus on how to meet various timing constraints as required in coverbal gesture. In summary, the problem of synchronizing synthetic gestural and spoken utterances has not been solved so far besides bringing single points in both modalities to coincidence. As commonly agreed upon, this is insufficient for virtual agents that shall be able to produce more extensive multimodal utterances in a smooth and life-like fashion.

3. Specification of utterances

Our approach to synthesizing multimodal utterances starts from straightforward descriptions of their overt form in a XML-based specification language (see Fig. 2). Such

descriptions contain the verbal utterance, augmented with nonverbal behaviors including gesture.

```
<definiton><utterance>
  <specification>
    And now take <time id="t1"/> this bar <time id="t2" chunkborder=
      "true"/> and make it <time id="t3"/> this big.<time id="t4"/>
  </specification>
  <behaviorspec id="gesture_1"> <gesture>
    <affiliate onset="t1" end="t2"/>
    <function name="refer_to_loc">
      <param name="refloc" value="$Loc-Bar.1"/>
    </function>
  </gesture> </behaviorspec>
  <behaviorspec id="gesture_2"> <gesture>
    <affiliate onset="t3" end="t4"/>
    <constraints>
      <symmetrical dominant="right_arm" symmetry="SymMS">
        <parallel>
          <static slot="HandShape" value="BSflat(FBround all o)
            (ThCpart o)"/>
          <static slot="PalmOrientation" value="PalmL"
            mode="absolute"/>
          <static slot="ExtFingerOrientation" value="DirA"/>
          <dynamic slot="HandLocation">
            <segment type="linear">
              <value type="start" name="LocShoulder
                LocCenterRight LocNorm"/>
              <value type="direction" name="DirR"/>
              <value type="distance" name="125.0"/>
            </segment>
          </dynamic>
        </parallel> </symmetrical> </constraints>
      </gesture> </behaviorspec>
    </utterance></definiton>
```

Figure 2. Sample XML specification of a multimodal utterance.

The correspondence between gesture and speech at this surface level is commonly assumed to exist between certain units on different levels of the hierarchical structure of both modalities [8, 13]. Kendon [8] defined *units* of gestural movement to consist of *gesture phrases* which comprise one or more subsequently performed movement *phases*, notably *preparation*, *holds*, *stroke*, and *retraction*. The intonational contour of connected speech in English and other languages (including German) is organized over *intonational phrases* (cf. [12]). Such phrases are separated by significant pauses and display a meaningful pitch contour with exactly one most prominent pitch accent (the *nucleus*)¹. We adopt the empirically suggested assumption that continuous speech and gesture are co-produced in successive units each expressing a single idea unit [13]. We define *chunks* of speech-gesture production to consist of an intonational phrase in overt speech and a co-expressive gesture phrase, i.e., complex utterances with multiple gestures are considered to consist of several chunks. Within each chunk, the gesture supports the conveyance of the prominent concept such that the stroke corresponds to the focussed constituent (the *affiliate*; a single word or subphrase of few words) in the intonational phrase, which has the nuclear accent [12].

¹If an element is prosodically focussed, the primary pitch accent expresses its prominence.

In our specification language, points in time are defined by inserting markups in the verbal utterance. Such time tags can be annotated to define chunk borders and, furthermore, are used to express cross-modal correspondence by specifying onset and end of the affiliate during the definition of a coverbal gesture (see Fig. 2). A particular gesture is stated in terms of the spatiotemporal features of its stroke, being the meaningful phase. The underlying idea of our XML-based gesture representation is that the gestural movement can be considered as a combination of submovements within three defining features: (1) the *location* of the wrist, (2) the *shape* of the hand, and (3) the *orientation* of the wrist (described by the direction of the extended fingers and palm orientation). Each feature value can be defined numerically or symbolically (using augmented sign language descriptions in *HamNoSys* [16]) and is constrained either to be held for a certain period of time (*static*) or to perform a significant submovement (*dynamic*) composed of subsequent *segments*. The overall structure of a gesture is given by the relationships between these feature constraints, e.g., moving the hand up while keeping a fist. To this end, *simultaneity*, *posteriority*, *repetition*, and *symmetry* of features can be denoted by specific XML elements constituting a constraint tree for the gesture. Instead of explicitly describing the features of an appropriate gesture, a desired communicative function of the behavior can be stated that is sufficient for the agent to choose a gesture from a lexicon of XML definitions.

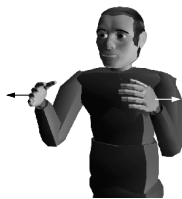


Figure 3. Iconic gesture as described in Fig. 2 (“gesture_2”).

Fig. 2 shows a description of an utterance comprising a deictic and an iconic gesture, the first one specified in terms of a required communicative function (“refer_to_loc”). The second gesture is defined by the movement of the right hand in addition to static wrist orientation and hand shape. Left hand movement results from the gesture’s mirror symmetry w.r.t. the sagittal plane (“SymMS”). The resulting utterance (in German language) created from this specification by our generation model (described in the following sections) can be found on the *Max* web page². The iconic gesture is depicted in Fig. 3. In addition to gestures, further nonverbal behaviors can be incorporated in the XML utterance spec-

²<http://www.techfak.uni-bielefeld.de/techfak/persons/skopp/max.html>

ifications (replacing the “gesture” tag). This includes arbitrary body movements and facial animations, defined as timed keyframe animations in joint angles or face muscle values.

4. Gesture planning and animation

The creation of gesture animations from feature-based descriptions requires a model for motion planning and control of upper limb movements. During higher-level planning (described in [10]), the given movement constraints of a gesture stroke are fully qualified (w.r.t. parameter assignment and timing), temporally ordered, and separately transferred to independent motor control modules. Currently, our system provides distinct modules for the hand, the wrist, the arm, as well as the neck and the face of the agent.

Since the integration of various motion generators is vital for simulating such complex hand-arm movements, we adopt a functional decomposition of motion control [21]: Each limb’s motion is kinematically controlled by a motor program which employs several *local motor programs* (LMPs) for animating necessary submovements within a limited set of DOFs and over a designated period of time (see Fig. 4). LMPs may differ in the employed motion generation method, working either in external coordinates or in joint angles. During lower-level planning, LMPs are instantiated and prepared by the motor control modules for their respective scope, e.g., the hand module adds LMPs affecting the joints of the hand. Since different movement phases may need to be created by different motion generators, LMPs can be concatenated by assigning predecessor and successor relationships. For example, the preparation and stroke phases in wrist bending are controlled externally by applying a quaternion interpolation LMP (“WristRot”), whereas wrist retraction is created in joint angle space.

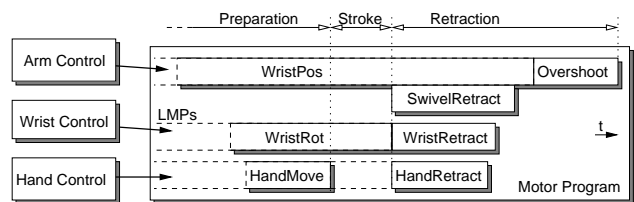


Figure 4. Composition of motion control for a hand-arm gesture.

In result, the motor program for a gesture stroke is created on-the-fly by the motor control modules from the feature constraints at their disposal. An example composition of low-level controllers is visualized in Fig. 4. However, since each gesture is planned separately, subsidiary gesture phases like fluent transitions can not be completely prepared

during planning. Therefore, LMPs act like behaviors, being able to activate themselves at run-time based on current movement conditions (indicated by the dashed edges of the LMP boxes in Fig. 4) and to connect themselves fluently to given boundary conditions. In addition, each LMP may transfer control between itself and its predecessor or successor, respectively. Once activated, LMPs are incessantly applied to the kinematic skeleton of the agent. For each frame, externally formulated LMPs (for wrist position and preparation/stroke of wrist flexion) are first invoked. Then, the inverse kinematics of the arm is solved using the analytical *IKAN* algorithm [19]. The arm’s redundancy is interpreted as “swivel” angle of the elbow about the shoulder-wrist axis and can either be controlled by a dedicated LMP or is determined using the sensorimotor transformation proposed by Soechting and Flanders [18] (also applied in the system of Koga et al. [9]). Finally, LMPs that directly affect joint angles (neck and hand motion, wrist retraction) can influence the solution configuration, e.g., by adding a continuous displacement or overriding a certain set of angles.

LMPs employ suitable motion generation methods. Of particular importance is the control of arm movement which can either be accomplished in joint angle space or in terms of the wrist trajectory in Cartesian coordinates. Since gestures have to reproduce shape properties as prescribed in the XML specification, we decided to create arm movement trajectories in working space, at least for the stroke phase. Next, we describe our method of constructing parametric curves that meet given position and timing constraints while resembling natural arm trajectories.

4.1. Formation of arm trajectories

Our approach to forming wrist trajectories rests upon the assumption that complex arm movements consist of subsequently and ballistically performed elementary units with the following stereotypical properties: (1) short targetted movements are approximately straight [14] and (2) exhibit a symmetrical bell-shaped velocity profile of the working point [11]; (3) a quasi-linear relation between amplitude and peak velocity, as well as an approximate logarithmic relation between amplitude and movement duration (*Fitts’ law*) holds. The segmentation of complex movements corresponds to points of maximum curvature of the working point trajectory (*breakpoints*) [14]. At this points, movement speed v drops and can be estimated from the radius r of the trajectory [11]: $v = b r^{\frac{1}{3}}$, where b is assumed to be constant within the space in front of the gesturer.

Relying on this assumptions, path and kinematic of arm movements can be reproduced based on the local behavior of the trajectory at the segmentation points. To this end, an intermediate representation is formed in the arm motor control module for each continuous movement phase, i.e.,

without any rest periods in between. It consists of a sequence of linear *guiding strokes* [14], each bridging from one segmentation point to the next, that can be combined to approximate curvilinear and circular segments. From this, a LMP is created (“WristPos” in Fig. 4) that, once it has activated itself at run-time, completes the trajectory formation by (1) inserting a preparatory guiding stroke from current start conditions, (2) estimating the velocities at interior segmentation points, and (3) constructing a parametric curve.

The direction of each velocity between two successive guiding strokes ($\vec{d}_{i-1} \rightarrow \vec{d}_i, \vec{d}_i \rightarrow \vec{d}_{i+1}$), given by the unit tangent \vec{n}_v , is assumed to be parallel to the chord through $\vec{d}_{i+1} - \vec{d}_{i-1}$. Since the velocity depends on the spans of the adjacent guiding strokes and the time intervals between their breakpoints u_j , the average velocity v is first estimated (eq. (1)). Since the reciprocal of the radius for small angles equals the rate of change of the tangent, $\frac{1}{r} = \|\dot{\vec{x}}(t)\|$, we assume that the tangent vector tends towards the direction of $(\vec{d}_{i+1} - \vec{d}_i)$, and estimate the trajectory radius (normalized to $[0 \dots 1]$) at u_i from the angle α between these two vectors (eq. (2)). Finally, the speed is derived by the “ $\frac{2}{3}$ -law”. Note that reducing the value of the constant b causes rather sharp bendings and lower velocities at breakpoints, whereas greater values lead to a more continuous and smoother movement. Yet, the constant must not be assigned a value too large to ensure that the velocity drops at the segmentation points.

$$v = \frac{1}{2} \left(\frac{\|\vec{d}_i - \vec{d}_{i-1}\|}{u_i - u_{i-1}} + \frac{\|\vec{d}_{i+1} - \vec{d}_i\|}{u_{i+1} - u_i} \right) \quad (1)$$

$$\alpha = \angle(\vec{n}_v, \vec{d}_{i+1} - \vec{d}_i), \quad r = 1 - \frac{\alpha}{\pi}, \quad v = b \cdot v \cdot r^{\frac{1}{3}} \quad (2)$$

Non-uniform cubic B-splines are applied to construct a composite curve which satisfies all position and velocity constraints at the breakpoints. Each interior knot is set to double multiplicity and equal to the corresponding breakpoint. This narrows the influence of the control points, intended from the fact that the movement should be determined by local properties at the breakpoints. Consequently, interior velocities become a major means of controlling the trajectory. The resulting C^1 -continuous spline, calculated from the position and velocity constraints, gives a smooth trajectory and reproduces symmetrical bell-shaped velocity profiles for each ballistical movement unit. Furthermore, the quasi-constant relationship between amplitude and maximum speed of human movement (for constant durations) is accounted for.

Dynamic properties may play an important role when verbal gestures are emphasized in synchrony with speech. Since accentuation of the movement can be created by an overall acceleration and additionally deferring the velocity peak (resulting in a shorter period of stronger deceleration), we extended our trajectory formation model by adjustable

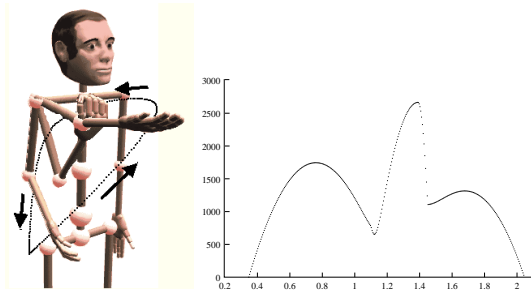


Figure 5. Sample arm movement and its speed profile (with deferred velocity peak).

placing of the moment of maximum velocity along single guiding strokes.

An example trajectory and velocity profile of a simple, fluently performed gesture (comprising three guiding strokes) are shown in Fig. 5. Bell-shaped speed profiles and lower velocity at the first breakpoint due to higher curvature are automatically reproduced. In addition, the velocity peak of the second guiding stroke (depicting a pulling action) has been deferred to emphasize the gesture’s execution. Further animation examples, all created in real-time solely from XML-based feature descriptions can be found at the aforementioned web page.

5. Speech-gesture coordination

The described gesture generation methods need to be embedded into a larger production framework that combines cross-modal synchrony within each chunk with a seamless flow of speech and gesture across successive chunks. To this end, planning and execution of subsequent chunks in our system are interleaved, each chunk being processed on a separate blackboard running through a series of processing states (see Fig. 6). During planning (*InPrep*), separate modules for speech synthesis and movement planning contribute concurrently to the overall planning process. In this stage, connecting effects are created when a following chunk is anticipated: The pitch level in speech is kept up and gesture retraction is planned to lead into an interim rest position. Once chunk planning has been completed, the state is set to *Pending*. If the chunk can be uttered, i.e., the preceding chunk is *Subsiding* (see below), the scheduler passes control over to the chunk (*Lurking*) whose behaviors (LMPs) activate themselves autonomously. Since until this point the preceding gesture may have not been fully retracted, fluent gesture transitions emerge depending on the placement of the affiliate within the verbal phrase (see Fig. 6). Depending on feedback information about all behaviors, which is permanently collected on the blackboard,

the chunk state then switches to *InExec* and eventually, once the intonational phrase and the gesture *stroke* have been completed, to *Subsiding* if some retracting LMPs are still active or to *Completed* otherwise.

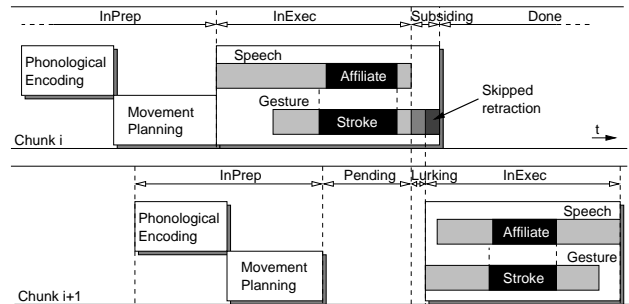


Figure 6. Incremental chunk processing.

In case the affiliate is located early in the intonational phrase, the gesture’s preparation precedes speech, i.e., the gesture “anticipates” the verbal utterance (as observed in humans [13]). Due to predefined movement acceleration limits, the vocal pause between subsequent intonational phrases may thus be stretched, depending on size and time-consumption of the preparation phase (cf. Fig. 6). Besides this adaptation to gesture, speech is articulated ballistically as prepared by the text-to-speech system.

Within each chunk, the gesture is scheduled such that the onset of the *stroke* precedes or coincides with the primary stress of the accompanying speech segment [13]. To this end, the assignment of a coverbal gesture in the XML specification affects the intonation of the verbal phrase. Utilizing a set of SABLE³ tags, the affiliate is marked to be emphasized by our text-to-speech system [20] that controls prosodic parameters like speech rate and intonation in order to create natural pitch accents. The resulting timing information is asserted to the chunk blackboard and utilized to compose lip-synchronous speech animations as well as to schedule the accompanying gesture: Applying *absolute-time-based* scheduling [5], the onset of the gesture stroke is set to precede the affiliate’s onset by approximately one syllable’s duration (0.2 sec). Moreover, the stroke is set to span the whole affiliate before retraction starts. For dynamic strokes, this is currently done by simply adjusting its execution velocity. Alternatively, the stroke may be performed in a comfortable speed and compensate for a longer affiliate with an extraordinary hold or additional repetitions (both strategies observable in humans [13]). Finally the gesture is dynamically created as described in Section 4.

³SABLE is an international standard for marking up text input to speech synthesizers.

6. Conclusion

Lifelike multimodal utterances of a great variety are highly desired for conversational agents. Instead of drawing predefined behaviors from fixed libraries as in most existing systems, all verbal and nonverbal utterances in our system are created on-the-fly from XML specifications of their overt form. A hierarchically organized model of motion control was developed for generating gesture animations in real-time. It comprises an informed method for forming a parametric curve that models natural path *and* kinematics for a required wrist movement in space. This novel approach to creating gesture animations from scratch provides satisfactory quality and allows to finely adapt the gestural movements to accompanying speech such that synchrony even at the phoneme level has been achieved. It is embedded in an overall framework for incrementally planning and executing complex utterances, which systematically employs co-articulation and transition effects in order to reconcile the continuous flow of speech and movement with temporal synchrony constraints. This exceeds the ability of current multimodal agents, in which synchronization of synthetic gestural and spoken utterances is accomplished by simply bringing single points of behaviors to coincidence, which are independent in any other respects. In future work, it seems natural to employ the flexibility of the presented method w.r.t. temporal adjustment of movement to enable a coordination of both modalities beyond the level of gesture stroke and affiliate. In particular, a synchronisation of the moments of stress in gesture and speech may yield to a coordinated accentuation, e.g., according to an underlying rhythmic pulse. This has to include the timing of velocity peaks of single movement phases, which has been already taken into account in our approach. Likewise, our text-to-speech system already offers mechanisms to synchronize stressed syllable with external events. In addition, the gesture animation model should be further explored w.r.t. variations of the constants that, e.g., influence the relationship between trajectory curvature and velocity. Modulating these values systematically within reasonable limits may enable the modulation of the agent's style of movement.

References

- [1] J. Cassell. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. In Cassell et al. [4], chapter 1, pages 1–27.
- [2] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. Human conversation as a system framework: Designing embodied conversational agents. In Cassell et al. [4], chapter 2, pages 29–63.
- [3] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH'94*, 1994.
- [4] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*. The MIT Press, Cambridge (MA), 2000.
- [5] J. Cassell, H. Vilhjalmsson, and T. Bickmore. BEAT: the behavior expression animation toolkit. In *Proceedings of SIGGRAPH 2001*, 2001.
- [6] E. Churchill, L. Cook, P. Hodgson, S. Prevost, and J. Sullivan. "May I help you?": Designing embodied conversational agent allies. In Cassell et al. [4], chapter 3, pages 64–94.
- [7] S. Gibet, T. Lebourque, and P. Marteau. High-level specification and animation of communicative gestures. *Journal of Visual Languages and Computing*, 12(6):657–687, 2001.
- [8] A. Kendon. Gesticulation and speech: Two aspects of the process of utterance. In M. Key, editor, *The Relationship of Verbal and Nonverbal Communication*, pages 207–227. The Hague, Mouton, 1980.
- [9] Y. Koga, K. Kondo, J. Kuffner, and J. Latombe. Planning motions with intentions. In *Proc. of the 21st annual Conference on Computer graphics*, pages 395–408, 1994.
- [10] 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 661–667, Amsterdam, 2000. IOS Press.
- [11] M. Latash. *Control of Human Movement*. Human Kinetics Publishers, Champaign, IL, USA, 1993.
- [12] W. Levelt. *Speaking*. MIT press, Cambridge, Massachusetts, 1989.
- [13] D. McNeill. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, Chicago, 1992.
- [14] P. Morasso. Trajectory formation. In P. Morasso and V. Tagliasco, editors, *Human Movement Understanding*. Elsevier Science Publishers B.V., North-Holland, 1986.
- [15] K. Perlin and A. Goldberg. Improv: A system for scripting interactive actors in virtual worlds. In *SIGGRAPH 96, Proceedings of the 23rd annual conference on Computer Graphics*, pages 205–216, 1996.
- [16] S. Prillwitz, R. Leven, H. Zienert, T. Hamke, and J. Henning. *HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide*, volume 5 of *International Studies on Sign Language and Communication of the Deaf*. Signum Press, Hamburg, Germany, 1989.
- [17] J. Rickel and W. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382, 1999.
- [18] J. Soechting and M. Flanders. Sensorimotor representations for pointing to targets in three dimensional space. *Journal of Neurophysiology*, 62(2):582–594, 1989.
- [19] D. Tolani, A. Goswami, and N. Badler. Real-time inverse kinematics techniques for anthropomorphic limbs. *Graphical Models*, 62:353–388, 2000.
- [20] I. Wachsmuth and S. Kopp. Lifelike gesture synthesis and timing for conversational agents. In *Proceedings of Gesture Workshop 2001*, London, UK, 2002. to appear.
- [21] D. Zeltzer. Motor control techniques for figure animation. *IEEE Computer Graphics & Applications*, 2(9):53–59, November 1982.