

Imitation games with an artificial agent: From mimicking to understanding shape-related iconic gestures

Stefan Kopp, Timo Sowa, and Ipke Wachsmuth

Artificial Intelligence Group
Faculty of Technology, University of Bielefeld
D-33594 Bielefeld, Germany
{skopp,tsowa,ipke}@techfak.uni-bielefeld.de

Abstract. We describe an anthropomorphic agent that is engaged in an imitation game with the human user. In imitating natural gestures demonstrated by the user, the agent brings together gesture recognition and synthesis on two levels of representation. On the mimicking level, the essential form features of the meaning-bearing gesture phase (stroke) are extracted and reproduced by the agent. Meaning-based imitation requires extracting the semantic content of such gestures and re-expressing it with possibly alternative gestural forms. Based on a compositional semantics for shape-related iconic gestures, we present first steps towards this higher-level gesture imitation in a restricted domain.

1 Introduction

Intuitive and natural communication with a computer is a primary research goal in human-computer interaction. This vision includes the usage of all communicative modalities, e.g., speech, gesture, gaze, and intonation, and has led to extensive work on processing a user's multimodal input as well as on creating natural utterances with humanoid agents. To address these problems with regard to gesture, we employ a scenario in which the human user is engaged in an imitation game with an anthropomorphic agent, *Max*. The human user meets Max in a virtual environment where he is visualized in human size. The agent's task is to immediately *imitate* any gesture that has been demonstrated by the user.

Imitation tasks are of particular interest as this capability can be considered a key competence for communicative behavior in artificial agents. Just like in infant development of communicative skills, two important steps towards this competence would be the following: First, the agent's capability to perceive the various behaviors in his opposite's utterance and to mimic them in a consistent way (*mimicking-level*). Second, and higher-level, to understand the meaning of the perceived utterance and re-express its content with his own communicative means, i.e., in his own words and with his own gestures (*meaning-level*).

Our previous work was directed to processing multimodal utterances of the user and to synthesizing multimodal responses of Max, both including coverbal

gestures [13, 7]. This contribution describes how the developed systems can be combined to enable the gesture imitation game on both aforementioned levels (see Fig. 1 for an overview):

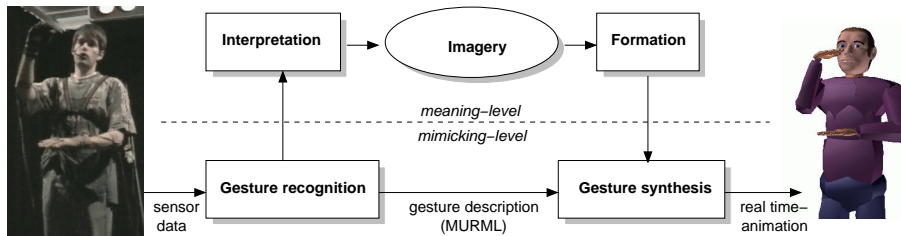


Fig. 1. Levels of gesture imitation (bottom-up: form-based and meaning-based)

On the mimicking level, Max employs a body-centered representation of the gesture as an interface between his recognition and production systems. Imitation therefore includes formally describing a gesture’s meaningful (stroke) phase in terms of its mandatory spatiotemporal features which transcends the direct transfer of low-level body configuration parameters like in motion capture. This kind of gesture imitation has been successfully realized for a certain range of gestures and is described in more detail in Section 3. Beyond mimicking, a recognized gesture can be conceptualized on the meaning level yielding a modality-independent representation of its content (idea unit). From this representation a gestural imitation can be generated that preserves the original communicative idea but may very well result in different realizations, e.g., dependent on the availability of expressive resources. For a domain limited to object shape descriptions by iconic gestures, we present a spatial representation for a gesture’s imaginal content in Section 4. First steps towards automatically building such representations, i.e., interpreting gestures, as well as forming particular gestures for them are described.

2 Related work

In spite of its significance for the cognitively plausible development of human-computer communication, imitation scenarios have not been adopted in previous systems and especially not so with respect to gestures. There is a large body of research on gesture recognition viewed as pattern classification. Systems based on this paradigm segment the input data stream from the sensing device and consider gestures as atomic pieces of information that can be mapped one-to-one on some application-dependent meaning. Popular computer vision methods include training-based approaches like Hidden Markov Models and Artificial Neural Networks, as well as explicit feature-based methods [16]. Multimodal approaches that include gesture and speech additionally consider the context-dependency of gesture semantics. Usually, some form of multimodal grammar is employed

to unify gesture and speech tokens in a common complex of meaning [4]. A semantic aspect which is particularly important in natural gestures is iconicity. However, most current systems do not provide any means to model gestural images explicitly based on their inner structure. One noteworthy exception is the ICONIC system that maps object descriptions with coverbal gestures on possible referents based on an explicit comparison of form features and their spatial configuration [15].

Similar to most recognition approaches, gesture generation in conversational agents (e.g. [12, 1]) usually relies on a fixed one-to-one mapping from communicative intent to predefined animations that are drawn from static libraries on the basis of specific heuristics. Although the animations can be parametrized to a certain extent or concatenated to form more complex movements, this approach obviously does not resemble a real “transformation” of meaning into gesture. Cassell et al. [2] present a system for planning multimodal utterances from a grammar which describes coverbal gestures declaratively in terms of their discourse function, semantics, and synchrony with speech. However, gesture production again does not associate semantic features with particular gesture features (i.e., handshape, orientation, movement) that would constitute a literally context-dependent gesture (cf. [2, p. 175]). A fully automatic gesture creation was targeted by only few researchers. Gibet et al. [3] apply generic error-correcting controllers for generating sign language from script-like notations. Matarić et al. [10] stress the problem of determining appropriate control strategies and propose the combined application of different controllers for simulating upper limb movements. For Max, we emphasize the accurate and reliable reproduction of spatiotemporal gesture properties. To this end, motor control is planned directly from required form properties and realized by means of model-based animation.

3 Mimicking: Form-based imitation

By gesture mimicking we mean the reproduction of the *essential* form properties of a demonstrated gesture by an articulated agent. This kind of imitation should be independent from the agent’s body properties (e.g., measures, proportions, level of articulation) and, furthermore, should not need to take subsidiary movement features into account. As depicted in Fig. 1, mimicking therefore includes (1) recognizing gestural movements from sensory input, (2) extracting form features of the gesture stroke and specifying them in relation to the gester’s body, and (3) synthesizing a complete gesture animation that reproduces these features in its stroke phase. This section describes the methods developed for all three stages and their combination in Max in order to enable real-time gesture mimicking.

3.1 Feature-based representation of gesture form

A gesture’s form features are described by a subset of MURML, a markup language for specifying multimodal utterances for communicative agents [7].

MURML defines a hand/arm configuration in terms of three features: (1) the location of the wrist, (2) the shape of the hand, and (3) the orientation of the wrist, compositionally described by the extended finger orientation/direction and the normal vector of the palm (palm orientation). Feature values (except for handshape) can be defined either numerically or symbolically using augmented HamNoSys [11] descriptions. Handshape is compositionally described by the overall handshape and additional symbols denoting the flexion of various fingers.

A gesture is described in MURML by specifying its stroke phase that is considered as an arbitrarily complex combination of sub-movements within the three features, e.g., moving the hand up while keeping a fist. To state the relationships between such features, simultaneity, posteriority, repetition, and symmetry of sub-movements can be denoted by specific MURML elements constituting a constraint tree for the gesture (see Figure 2 for an example).

```

<constraints>
  <parallel>
    <symmetrical dominant="right_arm" symmetry="SymMST"
      center="0 0 0 0 0 15.0" >
      <parallel>
        <static slot="HandShape" value="BSflat" />
        <static slot="ExtFingerOrientation" value="DirAL" />
        <static slot="HandLocation" value="LocChin
          LocCCenterRight LocNorm" />
      </parallel>
    </symmetrical>
    <static slot="PalmOrientation" value="DirD" scope="left_arm" />
    <static slot="PalmOrientation" value="DirD" scope="right_arm" />
  </parallel>
</constraints>

```




Fig. 2. Example specification of a static two-handed, symmetrical gesture

Each feature is defined over a certain period of time to be either (1) static, i.e., a postural feature held before retraction, or (2) dynamic, i.e., a significant sub-movement fluently connected with adjacent movement phases. For complex trajectories, dynamic constraints are made up of segments, which can be further differentiated for hand location constraints either as linear or curvilinear.

Like in HamNoSys, symmetric two-handed gestures are defined in terms of the movement of the dominant hand and the type of symmetry obeyed by the following hand. We define eight different symmetries made up of combinations of mirror symmetries w.r.t. the frontal, transversal, and sagittal body planes (“SymMST” in Figure 2 denotes mirror symmetries w.r.t. the sagittal and the transversal plane). Regardless of the symmetry condition, handshapes are identical in both hands and the wrist orientation vectors are always switched in transversal direction (left-right). Exceptions from this rule can be described explicitly by combining respective movement constraints with the symmetrical node (as in Figure 2 for palm orientation).

3.2 Gesture recognition and feature extraction

The gesture recognition stage transforms sensory data into a MURML form representation (Fig. 3). We employ a 6DOF tracking system and data-gloves to capture hand motion as well as posture information. Data from the devices is processed by a chain of software modules which are tightly integrated in the immersive virtual reality environment [9]. The modules compute form and movement features of MURML as well as cues for gesture segmentation. A specialized segmentation module processes such cues to divide the data stream into gesture phases based on the approach by Kita et al. [5]. A movement interval, including hand motion and posture changes, is segmented at points where an abrupt change of the movement direction occurs and the velocity profile significantly changes. An alternation of direction includes the transition between movement and holds. For each phase, a form description frame is created that includes the relevant features. The resulting frame sequence is searched for typical profiles that indicate a gesture phrase, for example, a phase in which the hand rises, followed by a hold, and then a lowering phase, is regarded as a *preparation-stroke-retraction* phrase. The hold in the middle would then be tagged as the meaningful phase and encoded in MURML.

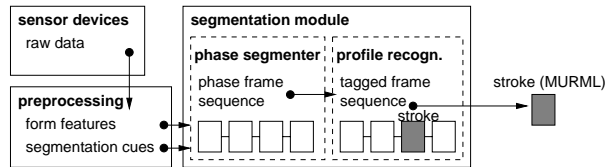


Fig. 3. Stages of the recognition process

3.3 Gesture synthesis

On the gesture synthesis stage, Max is able to create and execute gesture animations from MURML descriptions in real-time. An underlying kinematic skeleton for the agent was defined comprising 103 DOF in 57 joints, all subject to realistic joint limits. This articulated body is driven by a hierarchical gesture generation model, shown in Figure 4, that includes two stages: (1) high-level gesture planning and (2) motor planning. During gesture planning (see [6]), the expressive phase of a gesture is defined by setting up a fully qualified set of movement constraints. This stage optionally includes selecting an abstract gesture template from a lexicon, allocating body parts, expanding two-handed symmetrical gestures, resolving deictic references, and defining the timing of the stroke phase. During lower-level motor planning (described in [7]), a solution is sought to control movements of the agent's upper limbs that satisfy the constraints at disposal. A kinematic model of human hand-arm movement is employed that is based on findings from human movement science and neurophysiology. Each limb's motion is kinematically controlled by a motor program that concurrently employs low-level controllers (local motor programs; LMPs). LMPs animate sub-movements,

i.e., within a limited set of DOFs and over a designated period of time, by employing suitable motion generation methods. During planning, specialized motor control modules instantiate and prepare LMPs and subjoin them to the motor program(s). Our system provides modules for the hand, the wrist, the arm, as well as the neck and the face of the agent. Depending on body feedback during execution, blending of single gesture phases emerges from self-activation of the LMPs as well as the transfer of activation between them and their predecessors or successors, respectively.

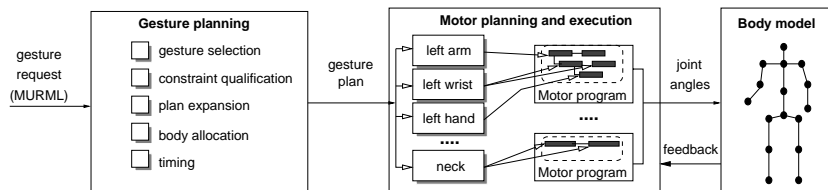


Fig. 4. Overview of the gesture generation model

Gesture recognition and synthesis have been connected by way of transferring the essential form features of a gesture encoded in MURML. That way, Max is able to mimic gestures in real-time standing face-to-face to the user (see Fig. 5 for examples). The recognition capabilities of the agent are currently limited to meaningful hold phases or turning points combined with arbitrary handshapes (e.g., as in pointing gestures). Dynamic movements and especially curvilinear trajectories pose particular problems for gesture segmentation as well as feature extraction and are subject of ongoing work.

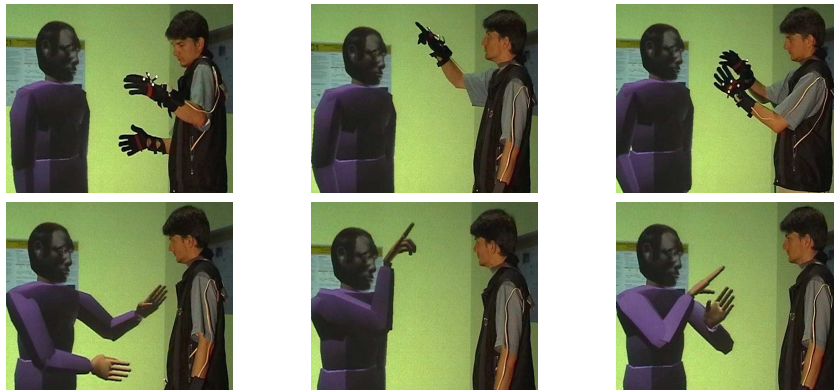


Fig. 5. Form-based gesture imitation: The form of a demonstrated gesture (upper picture) is mimicked by Max (lower picture)

4 Understanding: Meaning-based imitation

The recognition and generation systems described so far realize a first-level abstraction of upper limb movements to their essential form features (cf. Fig. 1). On this basis, we can approach meaning-based gesture imitation by deriving the idea unit behind the gesture from its form features (interpretation), formally representing this semantics, and re-expressing it in possibly alternative gestures (formation). To this end, plausible “sense-preserving” transformations between gesture morphology and meaning are needed. This section, first, describes a spatial representation for the semantics of iconic gestures for a domain limited to object shape descriptions. Experimentally derived, heuristic rules are presented for the mapping between this representation of gesture meaning and the feature-based descriptions of gesture form. Finally, first steps towards implementing this mapping in both directions, i.e., interpreting as well as forming shape-related iconic gestures, are described.

4.1 Imagistic representation of object shape

We conducted an empirical study to unveil the semantic features of shape and to determine the mapping between form and meaning of gestures in an object description task [14]. We observed that most gestures reflect an abstract image of their object which represents its extent in different spatial dimensions. We introduced the term *dimensional gestures* for this particular class. Such gestures are often reduced to convey just one or two dimensions, a phenomenon that we call *dimensional underspecification*. Following the object’s structure, gestural descriptions mostly decompose objects into simple geometric shapes that are described successively. Sometimes such descriptions contain an abstract sketch of the whole object before going into details. Successive dimensional gestures tend to coherently retain the spatial relations between the objects they represent. Groups of them may form complex elusive images in gesture space that reflect the qualitative arrangement of objects and their parts. For example, the main body of a riddled bar introduced by an iconic gesture may serve as a frame of reference for gestures indicating the position of the holes. Therefore, the spatial representation should cover larger semantic units spanning several gestures to allow for analyzing or synthesizing spatial coherence. We have chosen a structured, 3D representation called *imagistic description tree (IDT)* to model these properties. Each node represents an object’s or part’s basic spatial proportions. The approach is derived from a semantic representation schema for dimensional adjectives [8].¹ An *object schema* consists of up to three orthogonal axes describing the object’s extent. An *integrated axis* covers more than one dimension. A 2D-integrated axis can be regarded as the diameter of some object with a roundish cross-cut, and a 3D-integrated axis as the diameter of a sphere.

¹ The original implementation as a graph structure [13] allowed an arbitrary number of axes and spatial relations. It was modified to be compatible with Lang’s more restricted object schemas [8] which simplifies the recognition process.

Axis proportions and "meanings" can be qualitatively specified by the following attributes:

- max** defines the perceptually most salient axis which is commonly associated with the object's length. It is always disintegrated.
- sub** defines an integrated or disintegrated axis which is perceptually less salient. It is associated with the object's thickness, its material, or substance.
- dist** defines an interior extent, e.g., of non-material "parts" like holes.

Each axis' extent may be specified quantitatively with a length. To integrate verbal information, each schema contains names for the part it defines, e.g. "head" and "shank" for a screw. The tree structure models a part-of relation together with the spatial relation of a child node relative to its parent specified by a homogeneous transformation matrix. Thus it is possible to represent decomposition and spatial coherence. Fig. 6 shows an IDT for a stylized screw. It contains an underspecified schema for the whole object as root node as well as part schemas for the "head", the "shank", and the "slot" as a part of the head. The letters *a*, *b*, and *c* mark the spatial dimensions. Integrated axes are marked with parentheses, e.g. (*b c*). Attributes and quantitative values are listed below the axes.

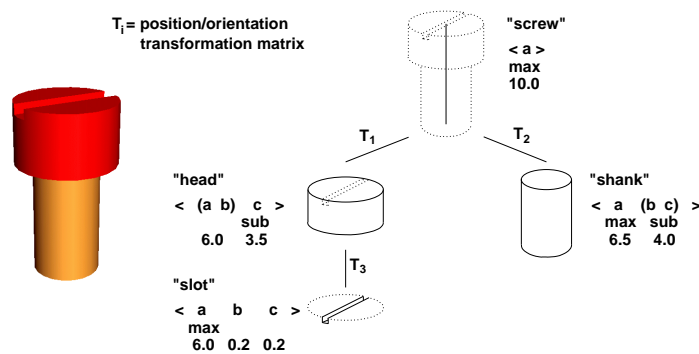


Fig. 6. Imagistic description tree for a stylized screw

4.2 The mapping between form and meaning

The gesture interpretation and formation stages concern the mutual conversion between a MURML form description and the semantic layer represented as an IDT. Indications for this mapping were obtained from the empirical study. It provides information about the form features used to gesturally encode spatial axes. Table 1 illustrates the combination frequencies of spatial features and form features in a corpus of 383 gestures judged as iconic. Note that the sum over all cells of the table exceeds 383 since several features may be present in a single gesture. The first column shows the number of gestural expressions of linear axes,

i.e. length, breadth, or width, the second one of the diameter of objects with a circular cross-cut. The third column lists cases in which a round shape, e.g. a hole, is indicated. The last two columns refer to indications of rounded corners or edges and the hexagonal shape of one of the stimulus objects. Axis, diameter, and round shape properties can be modeled with the imagistic description approach, whereas the model provides no means to describe the latter two. This weakness is acceptable since it affects only a small portion of the semantic object properties indicated by gestures (less than 3% of all descriptions).

	linear	diameter	round	r. edge/corner	hexagonal
movement	166	27	50	9	3
distance	87	40			
hand aperture	55	16			
palm orientation	15				
curved/round handshape			45		
index finger direction	1				

Table 1. Frequency of gesture form attributes expressing geometrical attributes

Generally, the relation between gesture form and meaning is a many-to-many mapping. However, Tab. 1 shows that a rather concise set of – still not unambiguous – heuristic rules can be derived if a compositional, feature-based approach is used on both the form- and the meaning-level. These rules are listed in Tab. 2.

#	form feature	axis type	axis orientation	axis degree
1	linear movement	disintegrated	orientation of movement segment	length of segment
2	circular movement	2D-integrated	movement plane	diameter of the circle
3	two-handed static posture, palms facing each other, flat handshape	disintegrated	difference segment between palms	distance between palms
4	two-handed static posture, palms facing each other, rounded handshape	2D-integrated	plane by difference vector between palms and extended finger orientation	distance between palms
5	precision-grip posture (thumb and other fingers facing each other)	disintegrated	vector between thumb tip and finger tip closest to thumb	hand aperture
6	flat hand (one-handed gesture)	disintegrated	extended finger orientation	(undefined)

Table 2. Rules for the conversion of form- and meaning-based representations

Complex images emerge either from the parallel use of features in a single gesture, or from their sequential expression in successive gestures. Since in the first case simultaneous form features must not occupy the same motor resources, the elementary mappings from Tab. 2 are composable in the following ways:

- 1 + 3: linear movement, two-handed symmetrical posture, flat hands
- 1+4: linear movement, two-handed symmetrical posture, rounded handshape
- 1 + 5: linear movement, precision-grip
- 3 + 5: two-handed symmetrical posture, precision-grip

Other combinations seem possible, but have not been observed in our corpus. Sequential arrangements of features, i.e., the distribution of an object schema across more than one gesture phrase, appear when incompatible form features are employed. An example is the description of a cube with three two-handed gestures of type 3, indicating successively its width, height, and length.

4.3 First implementation steps

The realization of meaning-based gesture imitation is an ongoing challenge in our lab. However, the imagistic shape representation provides an already operational basis for formalizing a gesture's imaginal content and the heuristic rules offer hints on how gestural form features can be associated. First steps towards utilizing these rules to automatically build an imagistic description tree from given gesture descriptions in MURML (interpretation) as well as to transform such a tree into MURML definitions (formation) have been taken. We expect that their combination will enable Max to recognize the image behind a sequence of iconic gestures and to create a different gestural depiction on his own.

Gesture interpretation For gesture interpretation the rules from Tab. 2 are basically implemented "from left to right". If a suitable form feature occurs in the MURML description, the corresponding axis type, its orientation and degree are inserted into the imagistic model. There are several possibilities for insertion depending on the feature arrangement strategy. Axes concurrently recognized, i.e. in a single dimensional gesture, always belong to one object schema. An axis expressed in sequence to another either completes the current object schema, or it opens up a new one. Furthermore, the algorithm has to decide where spatial coherence ends and a new gestural image begins. In its current state, the system assumes the beginning of a new image description if the hands return to a rest position after performing several gestures.

Gesture formation Starting from an imagistic description tree, gesture formation must cope with sequencing gestures that – in combination – are to express multiple object schemas with possibly multiple object features. Regarding this problem on a higher level, we assume that an object is described from its general spatial properties to more specific ones. This strategy resembles a depth-first

traversal of the imagistic description tree which, in addition, increases the spatial coherence in elaborating an object schema by describing its descendant schemas (e.g., the slot of the screw in Fig. 6 appears in relation to the head). For each object schema, a gestural expression is formed by iterating through its axes (from the dominant to the smaller ones), determining form features for each axis, and combining them if possible in accord to the aforementioned composition rules. Form features are selected according to the heuristic rules in Tab. 2. The ambiguity of this choice can be partially resolved based on the degree of the axis (e.g., feature 5 is selected for an axis of small degree rather than feature 1 or 3). In case feature selection is still ambiguous, the choice is done by chance. For each selected form feature, the corresponding movement constraints in MURML are created and adapted to the particular properties of the axis, e.g., hand aperture to axis degree. If the movement constraints of two or more axes cannot be combined due to conflicting consumption of body resources, separate gestures are instantiated as children of a sequence node in the MURML tree. All gestures formed for single object schemas are added into a single utterance, i.e. a single gesture unit, for the entire description tree. Currently, the composability of features of different axes is ignored during feature selection.

5 Conclusion

We have presented an approach for enabling artificial agents to imitate in a immediate, game-like fashion natural gestures demonstrated by a human user. The imitation scenario demands the connection of gesture recognition and synthesis methods. We propose two levels of gesture imitation, where representations of different degrees of abstraction are employed: On the mimicking-level, a gestural body movement of the user is reduced to the essential form features of its meaningful phase. These features proved successful for form-based gesture imitation when appropriate models are employed for both gesture recognition and synthesis. Gesture mimicking is demonstrable – so far limited to static gestures – in a real-time imitation game with Max. Building on the form-level abstraction, we further presented novel steps towards processing the meaning of iconic gestures that depict geometrical objects. A spatial representation for the semantics of such gestures was described along with experimentally derived rules that formalize the ambiguous mapping between form and meaning in an implementable way. Realizing this mapping in both directions is subject of ongoing work. In future work, we intend to include speech in the meaning-based imitation process. This would, for example allow the user to refer to an object using one modality (e.g., saying “bolt”) and getting it re-expressed by Max using the other modality (e.g., confirming the bolt’s shape with iconic gestures).

References

1. J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan. Embodiment on conversational interfaces: Rea. In *CHI'99 Conference Proceedings*, pages 520–527. ACM, 1999.

2. J. Cassell, M. Stone, and H. Yan. Coordination and context-dependence in the generation of embodied conversation. In *Proceedings of the International Natural Language Generation Conference*, pages 171–178, Mitzpe Ramon, Israel, June 2000.
3. 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.
4. M. Johnston. Multimodal unification-based grammars. Technical Report WS-98-09, AAAI Press, Menlo Park (CA), 1998.
5. S. Kita, I. van Gijn, and H. van der Hulst. Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth and M. Fröhlich, editors, *Gesture and Sign Language in Human-Computer Interaction: Proceedings of Gesture Workshop '97*, LNAI 1371, pages 23–36, Berlin, 1998. Springer-Verlag.
6. 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.
7. S. Kopp and I. Wachsmuth. Model-based animation of coverbal gesture. In *Proc. of Computer Animation 2002*, pages 252–257, Los Alamitos, CA, 2002. IEEE Computer Society Press.
8. E. Lang. The semantics of dimensional designation of spatial objects. In M. Bierwisch and E. Lang, editors, *Dimensional Adjectives: Grammatical Structure and Conceptual Interpretation*, pages 263–417. Springer, Berlin, Heidelberg, New York, 1989.
9. M. E. Latoschik. A general framework for multimodal interaction in virtual reality systems: ProSA. In *VR2001 workshop proceedings: The Future of VR and AR Interfaces: Multi-modal, Humanoid, Adaptive and Intelligent*, 2001. in press.
10. M. Mataric, V. Zordan, and M. Williamson. Making complex articulated agents dance. *Autonomous Agents and Multi-Agent Systems*, 2(1):23–44, July 1999.
11. 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.
12. 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.
13. T. Sowa and I. Wachsmuth. Interpretation of shape-related iconic gestures in virtual environments. In I. Wachsmuth and T. Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, LNAI 2298, pages 21–33, Berlin, 2002. Springer-Verlag.
14. T. Sowa and I. Wachsmuth. Coverbal Iconic Gestures for Object Descriptions in Virtual Environments: An Empirical Study. In M. Rector, I. Poggi, and N. Trigo, editors, *Proceedings of the Conference "Gestures. Meaning and Use."*, pages 365–376, Porto, Portugal, 2003. Edições Universidade Fernando Pessoa.
15. C. J. Sparrell and D. B. Koons. Interpretation of coverbal depictive gestures. In *AAAI Spring Symposium Series: Intelligent Multi-Media Multi-Modal Systems*, pages 8–12. Stanford University, March 1994.
16. Y. Wu and T. S. Huang. Vision-based gesture recognition: A review. In A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil, editors, *Gesture-Based Communication in Human-Computer Interaction (Proceedings of the Gesture Workshop 1999)*, Lecture Notes in Artificial Intelligence (1739), pages 103–115, Berlin, 1999. Springer-Verlag.