Coverbal Iconic Gestures for Object Descriptions in Virtual Environments: An Empirical Study

Timo Sowa & Ipke Wachsmuth
University of Bielefeld

Abstract. This paper describes an empirical study aimed at investigating object references in Virtual Environments using iconic gestures. Observations are focused on spatial concepts conveyed gestually and their relation to features of the gesture shape. A set of important features and spatial concepts useful for automated gesture recognition is identified. Based on these findings we propose a model of an iconic reference recognizer.

1 Introduction

An important research area in human-computer interaction deals with the conception of interfaces enabling users to interact "naturally" with a computer. Whereas mouse- and keyboard-based interaction is still the standard, research progresses towards user-centered and intuitive interfaces for highly interactive tasks. In particular, applications that provide a Virtual Environment (VE) demand for new interaction styles that overcome limitations and inefficiencies of traditional input techniques. Commonly, a VE is understood as an artificial surrounding generated by means of computer graphics. The interaction styles in VEs generally are subject to two different metaphors. Most approaches so far favor the metaphor of direct manipulation in which interaction in the virtual space resembles physical acts in reality. A more recent metaphor offers the view of communication between the user and an intelligent mediator that manipulates the virtual scene according to the user's intentions. A communicative interface of this kind profits from the integration of natural language and gesture which constitute primary modalities of human communication. Our approach treats the latter case for the creation of gesture and speech interfaces to virtual construction and design applications. In our recent work, we concentrate on the recognition and interpretation of iconic gestures *(iconics)* that provide a suitable way to communicate spatial properties of virtual entities.

In this paper, we present an empirical study investigating the use of iconic gestures for object descriptions in VEs and draw conclusions with respect to the design of a gesture interpretation system. The main questions to be answered are "Which spatial concepts are apparently expressed in iconic gestures?", "What elements of the gestures convey these concepts?", and "How can we implement this knowledge in a computer system?". Section 2 collects some important definitions of and findings about iconic gestures from the field of psycholinguistics and semiotics. Section 3 describes how multimodal interfaces, and especially iconic gestures, could be used in VE applications. An existing technical approach towards iconic gesture recognition is discussed in more detail. In Section 4 we present our empirical study, an evaluation and its analysis. Finally, a model of a gesture recognition system for iconic object references is proposed in Section 5.

2 Iconic Gestures

According to the gesture typology by McNeill (1992), iconics are pictorial gestures that "bear a close formal relationship to the semantic content of speech." This class of hand movements was first introduced by Efron (1941/1972) who called them physiographic gestures. They belong to the much broader category of illustrators in the classification schema by Ekman and Friesen

(1969). They define illustrators as “movements which are directly tied to speech, serving to illustrate what is being said verbally” (p. 68). The conception of illustrators is further subdivided in six categories, among which spatial movements, kinetographs and pictographs roughly resemble iconic gestures in the sense of McNeill. Rimé and Schiaratura (1991) give an overview of various naming conventions and definitions for iconic gestures. By emphasizing the close and direct coupling with speech, the above definitions implicitly show the semiotic characteristic of iconic signs in the sense of Peirce (1965), who distinguishes three types of signs: icons, indices and symbols. Unlike words and emblems, iconic gestures are innately related to their referent by similarity. As Peirce coins it, an icon “is a representamen which fulfills the function of a representamen by virtue of a character which it possesses in itself, and would possess just the same though its object did not exist” (5.73). An iconic movement or hand posture thus represents its object (meaning) purely by “existence”, rather than by shared cultural background or social agreement. However, iconicity should not lead to the misconception that a perfect image is drawn in gesture space using always the same gesture shape. There may be differences between individuals in establishing similarity, for example, people may use different limbs to express the same object iconically. Feyereisen and de Lannoy (1991, p. 11) point out that iconicity may “not relate to the relationship between the sign and its referent but to the relationship between the sign and its ground”, that is “the aspect, by which the sign evokes the referent”. Thus, the “iconic function”, that maps from gesture to object, is not a bijective one.

We believe that iconic gestures are a promising way to improve human-computer interfaces in spatial domains. This view presupposes a communicative value of gestures, an assumption on which we will comment briefly, since this topic is still discussed quite controversial. The more traditional standpoint taken by Kendon (1994), McNeill (1992) and others emphasizes the primacy of communication over other functional roles. Thus – as Kendon (1985, p. 27) puts it – “gesticulation arises as an integral part of an individual’s communicative effort”. Beattie and Shovelton (1999) found empirical evidence supporting this claim for iconic gestures. An alternative explanation is offered by other authors, for instance Krauss and Hadar (1999), who support the view that iconics are facilitators for lexical retrieval. They argue that “in order for a linguistic message to be communicative, the speaker must ... intend the message to create some particular effect ... in the addressee” (p. 94). These authors doubt that iconic gestures are intentional utterances and conclude that their communicativeness is at least questionable. However, in our view the question of intentionality is a minor issue, because even when a gesture was uttered unintentionally, it may contain valuable information that helps the addressee (i.e. the computer in our case) to understand.

The determination of meaning for iconics is quite difficult, it may even be impossible without further information. Due to the tight semantic coupling of iconic gestures and speech, verbal utterances seem necessary for a gesture to be understandable (Hadar and Butterworth, 1997). In fact, Feyereisen et al. (1988) have shown that people perform very poorly, when trying to indicate the meanings of gestures from videotapes without sound. Though speech appears to be the primary frame of reference for an iconic gesture, it is not the only one. The entire context of the situation in which the gesture is performed contributes to the meaning. The strong emphasis of speech may be a side-effect of narrative discourse analysis, being the dominant scenario for many efforts on gesture research in the past. In narrative discourse the world around is more or less irrelevant since the speaker establishes “common ground” by means of the narration. But this is not the case if we talk about and interact in the current situation. Thus, we prefer to say that the context of the situation is necessary for an iconic gesture to be understandable.

While context-free interpretation is not viable, it seems nevertheless possible to differentiate meaningful hand movements, including iconic gestures, from other movements, even for untrained persons (Feyereisen et al., 1988). This observation suggests that physical features exist which are characteristic for meaningful movements. McNeill (1992) subdivided the gesture space in front of the body into different regions and showed that there is a tendency to perform iconic gestures in the central region, whereas other types of gestures, like deictics and beats, can be found in more peripheral space. Hadar (1989) proposed three outstanding features defining iconic gestures: (1) They are complex, compared to the number of vectorial components needed for a description, (2) they have a wide amplitude (unlike beats) and (3) a relatively long duration. With respect to syntagmatic structure, gestures are generally characterized by successive movement phases. Kendon (1980) subdivided a “G-Unit” (a single gesture) in preparation, stroke, hold, and recovery phase. An additional pre-stroke hold after the preparation phase was observed by Kita (1990). He suggested that holds are used for synchronization with concurrent speech. In the preparation phase, the hands are brought into position, the obligatory stroke phase conveys the meaning, and the recovery phase moves the hands back to a resting position. Kita et al. (1998) found reasonable agreement between different human coders with respect to the transcription of movement phases. However, this kind of “gesture syntax” is not a strict one, since all phases except the stroke may be omitted.

Having said all this, we are inclined to draw the following conclusions:

(1) Even though iconic gestures are seldom found without accompanying speech, the presence of an iconic gesture can often be determined solely by significant features observable in body movement and independent of an evaluation of the language transmitted.

(2) While most studies have investigated iconic gestures in the context of narrative discourse, it is our observation that iconics may also be understood in a situational context, sometimes without language.

(3) Even when iconic gestures may often be produced unwillingly, as some authors have argued, they nevertheless occur as part of a multimodal utterance by which the utterer intended to communicate a quality to an addressee.

Hence we believe iconic gestures to be in a causal relation with an intended utterance and thus a means of communication. This is our starting point to investigate iconics in the context of human-machine communication as we do in the following sections.

3 Iconic Gestures in Virtual Environments

3.1 VE Applications and Gesture Input

Virtual construction and design are our central application areas of multimodal, i.e. gesture and speech-driven interfaces. In our laboratory, the system user stands in front of a wall-sized display and interacts with the system using gestures and speech. Concrete tasks may include virtual, knowledge-supported assembly of product prototypes from unit construction systems or interior design tasks (Wachsmuth & Cao, 1995; Latoschik et al., 1998). By combining virtual reality visualization techniques and multimodal interfaces, objects and designs become imaginable and the applications are easier to use. Interactions of this kind typically involve spatial references, e.g. the selection of an object or the manipulation of spatial relations and entities. The recognition and interpretation of instructions including iconic gestures is assumed to be an efficient way to communicate such information, since iconics are directly linked to the spatial domain.

In contrast to deictic and symbolic gestures which have been subject to many efforts in human-computer interaction (Wachsmuth & Fröhlich, 1998), iconic gestures have so far found little attention in technical settings. Apart from approaches that exploit some sort of iconics in pen-based interfaces (Cohen et al., 1997), the ICONIC system so far seems the only effort to make use of iconic gestures in virtual environments. It has shown to some extent how iconic gestures can be made workable as a means in human-machine communication, yet there is ample space for further advancements as we describe in more detail below.

3.2 An earlier Approach towards Iconic Gesture Interpretation

The ICONIC system was presented by Sparrell and Koons (1994); see also (Koons et al., 1993). Their prototype "allows a user to interact with objects in a computer generated environment through free speech and depictive gestures" in order to place and move them in the virtual scene (p. 10). The interpretation of an utterance is described as a three-step process: First, raw data is transformed to a gesture feature representation containing hand-shape, position and movement data. In a second step, the system considers gesture phase and timing. Whenever speech suggested the possibility of a gesture, the system searched for an appropriate stroke segment that was close in time. The gesture segmentation was performed by means of the typical gesture phases ”preparation", "stroke" and "retraction", with pauses necessary before and after the stroke. Finally, the meaning was determined. Static gestures were used to indicate the places of objects, whereas dynamic gestures indicated movements. Additionally, hand postures were matched against the shape of objects for reference purposes. If a hand-shape resembles the object shape, orientation parameters of the hand were applied to the object. The mapping was done by comparison of basic hand postures and object components like corners, flat sides, major and minor axes or default directions.

Although the ICONIC approach explicitly takes iconicity into account and therefore overcomes a simple shape-to-meaning mapping (reducing gesture interpretation to a pattern-matching task), there are some questionable restrictions. First, gesture segmentation relies on clear-cut preparation, pre- and post-stroke hold and retraction phases. However, this prototypical schema is often violated in natural gestures by leaving out phases, since only the stroke is obligatory. A less rigid way to determine stroke phases would be desirable. A second weakness is the distinction between static and dynamic gesture interpretation, indicating places or movements, resp. This difference seems quite arbitrary, for dynamic gestures may also be used to indicate other aspects besides movements, as we found in our experimental corpus (see following section).

4 Empirical Study

4.1 Motivation

Linguistic and psychological research have already provided many studies about iconic gesture behavior. Nevertheless, the question *how* iconicity is accomplished is not yet examined in detail. Our study is a first step towards this question, however restricted to a description task in a technical environment aimed towards multimodal human-computer interfaces. Thus, we also have to consider the scenario in which interaction takes place. In contrast to an ordinary dialogue situation there is no human partner but a screen with virtual objects, and the user wears technical devices for posture and movement detection.

4.2 Experiment Design

A total of 37 subjects (all of them German native speakers) were asked to describe five virtual objects (referred to as object A, B, C, D and E) presented one by one on a wall-size display (Fig. 1). The objects were taken from a system that is used in a virtual construction application.

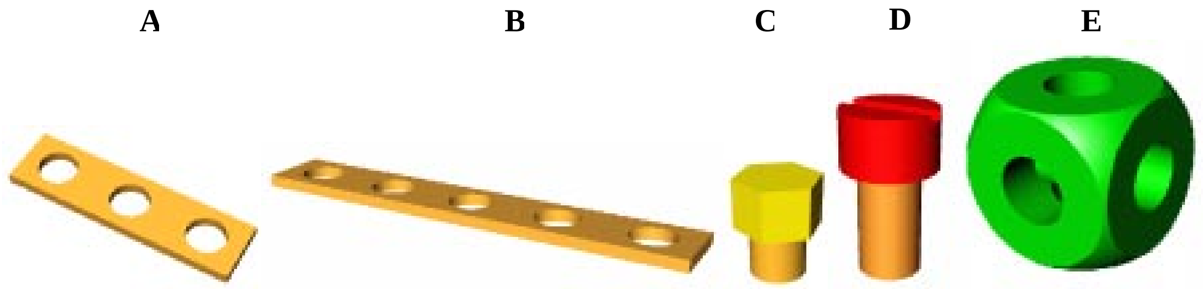

Figure 1: Virtual objects used as stimuli

The subjects wore a microphone, two data-gloves, and three trackers mounted at the wrists and the neck, to record speech and motion data and to make the results comparable with the application environment that we use in our lab (Fig. 2). The sensor devices were explained to the subjects; usage of gestural descriptions was not enforced, but suggested as an option if subjects inquired. No further constraints were imposed on the subjects. They were videotaped from a frontal view while they described the objects presented (Fig. 3).

4.3 Analysis Method

In a first analysis step, video recordings were chosen in which subjects performed iconic gestures. We accounted each movement an iconic gesture that obviously (as judged by one of the authors) referred to spatial properties of the object currently displayed. Fig. 3 shows an example of a subject indicating the size of a virtual object. It turned out that 25 subjects out of 37 (68%) used iconics in this sense. Subsequently, transcriptions of the verbal and gesture modalities were made for objects which were described using iconics. The words uttered were written down and annotated with movement phases (preparation, stroke, hold, retraction) of the co-occurring gestures. Every stroke phase was annotated with a posture description using the *HamNoSys* notation system (Prillwitz, 1989) and, in case of a dynamic gesture, a natural language movement description. Symmetries in the gesture shape were noted explicitly. The gesture interpretation, i.e. geometrical aspects of the object which resemble the gesture shape, was judged under consideration of the verbal channel. To give a rough idea, Figure 4 shows a sample transcription.

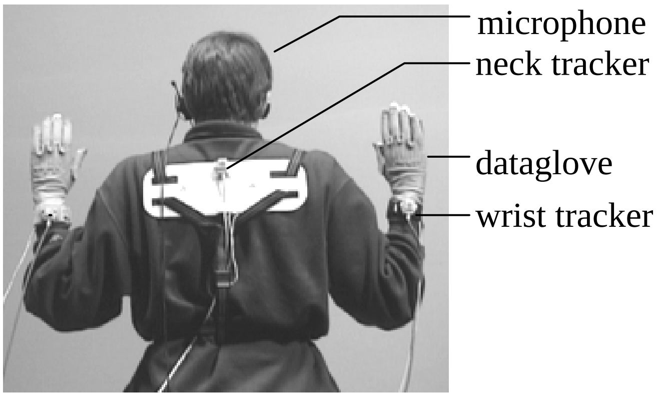

Figure 2: Sensor equipment

Figure 3: Example from the video recording

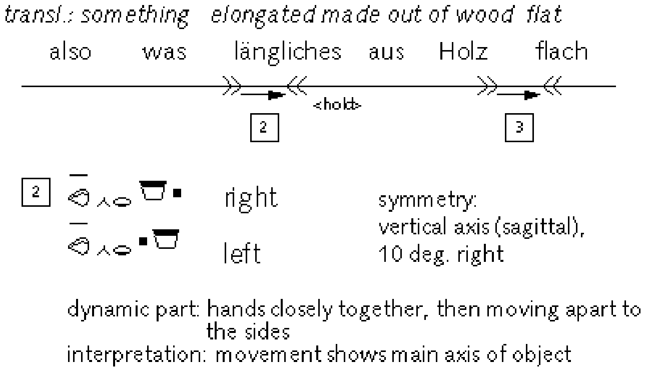
transl.: something elongated made out of wood flat
also was längliches aus Holz flach
<hold>
2
3
2
right
left
symmetry:
vertical axis (sagittal),
10 deg. right
dynamic part: hands closely together, then moving apart to
the sides
interpretation: movement shows main axis of object

Figure 4: Sample transcription using HamNoSys and verbal annotations

The HamNoSys part of the transcript contains a symbol string coding the postures of both hands. Each symbol denotes one of the aspects hand-shape, hand orientation, and hand position. The transcriptions for the 25 subjects that used gestures judged as iconic yield a corpus of 383 iconic gestures that were analyzed. The number of gestures used for each object, totalled across all subjects, is shown in Table 1. The lower gesture frequency for the long bar (Object B) may be explained by the similarity of stimuli A and B, which saved subjects some explanation effort. The higher frequency for the cube (Object E) may be an effect of its distinct three-dimensionality. Some subjects used combinations of gestures to show the proportions of the cube in all the three axes, leading to a higher gesture count. Gesture frequency for the other three objects is rather well balanced.

Object	A	B	C	D	E
No. of gestures	70	58	74	77	104

Table 1: Number of gestures per object across all subjects

4.4 Gesture Form Analysis

The corpus was classified according to general properties that roughly describe the gesture form. The categories *dynamic gesture, two-handed gesture* and *symmetrical gesture* were judged by a simple yes/no decision. Gestures are called dynamic if their stroke phase (i.e. the phase in which meaning is expressed) includes some sort of movement and they are called two-handed if both hands contribute to the expression of meaning. Two-handed gestures were classified as symmetrical if the positions or movements of both hands are either mirrored or parallel. Additionally, the number of *repetitions* in dynamic gestures was counted. Movements were considered to be repetitive if a first movement phase was followed by at least a second (reverse) phase, leading the limb back to its initial position or orientation. Hand-shapes were classified into six categories that were derived from the HamNoSys descriptions and that turned out to be characteristic features after a first corpus inspection: *flat, round, open precision grip, closed precision grip, stretched index, straddled fingers* (Fig. 5).

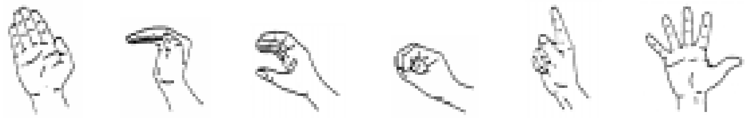

Figure 5: Hand-shape categories according to Prillwitz et al. (1989)

Finally we classified the movement types for dynamic gestures. We distinguished between *linear movements*, *circular/arc (-segment) movements* and *orientation change*, i.e. a stationary movement leaving the hand position unchanged. In more detail, Table 2 shows category frequencies for the whole corpus as well as for each of the five objects. Tables 3 and 4 show the frequencies of the hand-shape and movement categories that we observed.

	Total	*Object A*	*Object B*	*Object C*	*Object D*	*Object E*
Dynamic	249 (65%)	50 (71%)	41 (71%)	31 (42%)	54 (70%)	73 (70%)
Two-handed	211 (55%)	40 (57%)	32 (55%)	54 (73%)	33 (43%)	52 (50%)
Symmetrical	200 (52%)	39 (56%)	31 (53%)	52 (70%)	32 (42%)	46 (44%)
Repetitive	49 (13%)	5 (7%)	2 (3%)	13 (18%)	20 (26%)	9 (9%)

Table 2: General form categories and frequencies across all subjects

	Total	*Object A*	*Object B*	*Object C*	*Object D*	*Object E*
Flat	122 (32%)	17 (24%)	18 (31%)	23 (31%)	21 (27%)	43 (41%)
Round	37 (10%)	3 (4%)	2 (3%)	11 (15%)	8 (10%)	13 (13%)
Open prec. grip	95 (25%)	24 (34%)	19 (33%)	21 (28%)	23 (30%)	8 (8%)
Closed prec. grip	9 (2%)	3 (4%)	1 (2%)	1 (1%)	2 (3%)	2 (2%)
Stretched index	104 (27%)	21 (30%)	18 (31%)	10 (14%)	20 (26%)	35 (34%)
Straddled fingers	16 (4%)	2 (3%)	0	8 (11%)	3 (4%)	3 (3%)

Table 3: Hand-shape categories and frequencies across all subjects

	Total	*Object A*	*Object B*	*Object C*	*Object D*	*Object E*
Linear	172 (45%)	39 (56%)	32 (55%)	16 (22%)	36 (47%)	49 (47%)
Circular/Arc	60 (16%)	9 (13%)	6 (10%)	11 (15%)	13 (17%)	21 (20%)
Orientation change	16 (4%)	2 (3%)	2 (3%)	4 (5%)	5 (6%)	3 (3%)

Table 4: Movement categories and frequencies across all subjects

It is remarkable that the majority (65%) of gestures in our corpus were considered as dynamic although the stimuli were static images. This result invalidates the assumption of Sparrell and Koons (1994) who directly related gesture dynamics to referent dynamics (see Sect. 2.3). Most of the two-handed gestures (making up 55% of the corpus) are symmetrical. They amount to 52% of the corpus, this is a fraction of 95% of all two-handed gestures. Therefore, symmetry seems to be an important form feature that indicates meaningful hand movements. Our gesture corpus generally obeys the *symmetry condition* that was formulated for sign-languages by Battison (1978). It states that in case of independent hand movements both hands make the same or a symmetrical movement, have the same hand-shape, and the same or symmetrical orientation as well as the same location. Van Gijn et al. (1999) tested the condition for natural gestures and found evidence for a general applicability. Exceptions from the rule mostly referred to situations in which the hands had two separate target representations. This case rarely occurs in our corpus as stimuli did not contain more than one object at a time, which provides an explanation for the low number of non-symmetrical two-handed gestures. Thus, hand-shape and movement analysis for one hand was found to be sufficient. Repetitive movements emphasize the gesture stroke and may therefore be used as an additional cue to detect the meaningful part, though their frequency in our corpus is not particularly high. We observed up to seven repetitions for one gesture.

4.5 Meaning Categorization

A qualitative analysis of the corpus revealed that the gestures convey shape properties in various ways. First, they may constitute an abstraction from the complete object shape, reducing, for example, a bar to a one-dimensional line or a two-dimensional surface. Further, shape descriptions using gestures often reflect the spatial extension of an object with respect to certain axes or planes. A general observation pertains to the subdivision of single objects, i.e. their structural decomposition into parts described independently. Objects were always decomposed by the subjects, i.e. we observed no case in which the whole object shape was depicted in a single complex gesture. Fig. 6 shows the objects and their components, e.g. head and shank of the screws but also immaterial cognitive entities like holes and the screw slot, which were described by individual gestures.

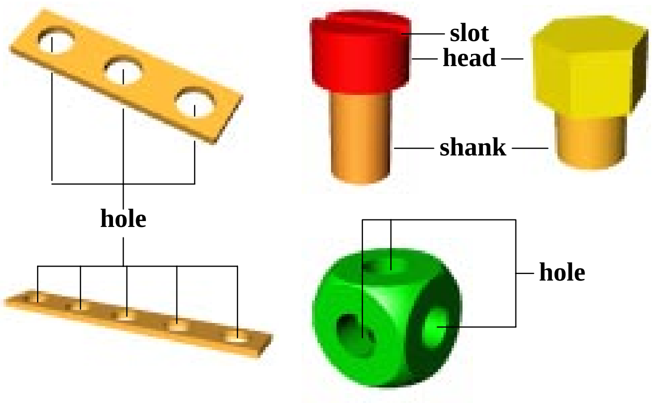

Figure 6: Object decomposition

To analyze the semantic contribution of gestures to the multimodal utterance, a set of meaning-related categories for each of the five objects was put up in a first analysis step. The categories were chosen to be in accordance with the observed properties of abstraction and decomposition. They are listed below; note that a single gesture may be classified in multiple categories.

Objects A, B:

a_1	primary extension of the bar (main axis, “length”)
a_2	secondary extension (“width”)
a_3	tertiary extension (“thickness”)
hole	shape of the holes

Objects C, D:

a_1	primary extension of the screw (main axis, “length”)
a_h	primary extension of the screw’s head
a_s	primary extension of the screw’s shank
r_s	round shape of the shank
d_s	secondary extension of the shank (diameter)
d_h	secondary extension of the head (diameter)
r_h	round shape of the head
h_h	6-sided shape of the head (used for object C only)
a_{s1}	primary extension of the screw’s slot (used for object D only)

Object E:

a_1	axis through left and right side (“width”)
a_2	axis through upper and lower side (“height”)

a_3	axis through frontal and back side (“depth”)
ha_1	axis of the hole from left to right
ha_2	axis of the hole from top to bottom
ha_3	axis of the hole from front to back
hole	shape of the holes
r_s	round shape of the side faces
$r_{c/e}$	rounded shape of corners/edges

In Table 5 frequencies of each category according to this categorization scheme are summarized for each object.

A	Attribute	a_1	a_2	a_3	*hole*
	Frequency	53 (76%)	27 (39%)	11 (16%)	11 (16%)

B	Attribute	a_1	a_2	a_3	*hole*
	Frequency	47 (81%)	13 (22%)	13 (22%)	8 (14%)

C	Attribute	a_1	a_h	a_s	r_s	d_s	d_h	r_h	h_h
	Frequency	4 (5%)	10 (14%)	18 (24%)	19 (26%)	25 (34%)	24 (32%)	9 (12%)	3 (4%)

D	Attribute	a_1	a_h	a_s	r_s	d_s	d_h	r_h	a_{sl}
	Frequency	3 (4%)	5 (6%)	20 (26%)	17 (22%)	18 (23%)	16 (21%)	13 (17%)	24 (31%)

E	Attribute	a_1	a_2	a_3	ha_1	ha_2	ha_3	*Hole*	r_s	$r_{c/e}$
	Frequency	18 (17%)	17 (16%)	4 (4%)	20 (19%)	11 (11%)	7 (7%)	14 (13%)	4 (4%)	10 (10%)

Table 5: Frequencies of object attributes across all subjects

4.6 The Relationship between Meaning and Form

With the set of meaning-related geometrical attributes introduced above it is now possible to analyze how these attributes are expressed by form features of the gesture. Table 6 shows the relevant form features that were judged to convey information about spatial extensions (axes), round shapes, rounded corners/edges and hexagonal shape. The form categories we chose are *movement trajectory, hand distance, hand aperture, palm orientation, curved/round hand-shape* and *index finger direction*. Again, it becomes apparent that movement conveys meaning even in our static scenario. Both linear movement, for axis indication, and circular movement, for indication of roundness, are found. Besides movement, *hand distance*, or, more precisely, the vector between left and right palm in symmetrical gestures, and *hand aperture*, i.e. the vector between the tip of the thumb and the other fingers, are often significant features. The same is true for curved or round hand-shapes, which are often used to indicate a curved or round object feature.

	a_n	d_n	r_n, *hole*	$r_{c/e}$	h_h
Movement	166 (51%)	27 (33%)	50 (53%)	9 (100%)	3 (100%)
Distance	87 (27%)	40 (48%)			
Hand aperture	55 (17%)	16 (19%)			
Palm orientation	15 (5%)				
Curved/round hand-shape			45 (47%)		
Index finger direction	1 (0.3%)				

Table 6: Frequency of gesture shape attributes expressing aspects of geometry

4.7 Quantitative Aspects

One further issue that was given attention is concerned with quantitative aspects of iconic gestures. When subjects indicate axes or other geometrical properties, is it just shape or do they also convey quantitative information like absolute size and orientation? To answer this question, sizes of gestures were compared with corresponding sizes of the objects as they appeared on the projection wall. A subset of gestures was chosen that apparently conveyed information about object extensions by hand distance. The large deviation between gesture and size suggests that it is not possible to map gesture sizes directly on referent sizes (see Table 7). A similar observation was made for object orientation, which is conveyed by some, but not all, gestures. Although no quantitative evaluation was made of this aspect, we feel that – even when subjects apparently meant to convey orientation by gestures – exploitation of such information would be very complicated, if possible at all.

True size (cm)	Gesture size (cm)
79	65, 65, 20, 52, 5, 65, 36, 43, 59, 71, 87, 20, 48
140	71, 87, 20, 48
81	22, 44, 53
96	44, 48, 45, 20, 12
80	14, 70, 73, 22, 44, 39

Table 7: Object size vs. size indicated in gestures

5 A Model for Iconic Gesture Interpretation

The ultimate aim of our research is to exploit iconic gestures for object references in Virtual Environments. Putting together the results from our corpus analysis, we propose a rough model for a gesture interpretation system to resolve such references. The model, as sketched in Fig. 7, consists of three computational modules – a feature recognizer, a spatial entity coder, an iconic mapper – and two memory modules storing gesture input and object knowledge.

In a first step, the sensor data that comes from the motion trackers and data gloves is to be analyzed and classified by a feature recognizer. The feature set should (at least) include the shape properties observed in our experimental corpus, i.e. hand-shapes (see Fig. 5), linear and curved movements, hand orientation changes, symmetries, and repetitions.

In a second step, gesture features are abstracted to basic spatial entities in a spatial entity coder. Our analysis suggests to start with bounded linear segments and circular segments, since these features were dominant in our corpus. The abstraction step neglects how certain entities were actually expressed by the gesture. The aim is to make it possible, for example, to identify both a

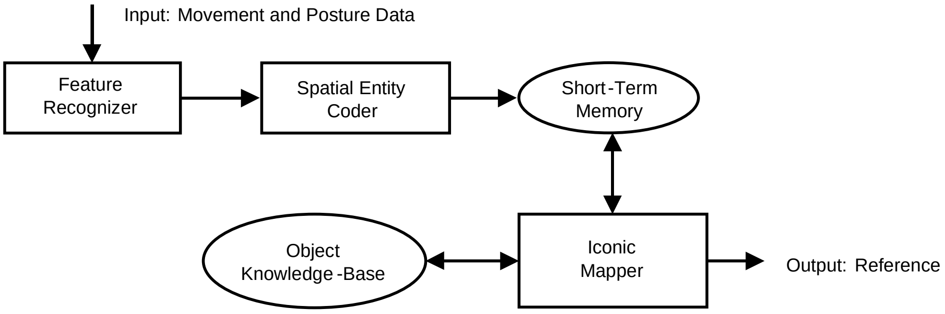

Figure 7: A model of a gesture recognizer for iconic references

linear movement of the hand and a two-handed static gesture with palms facing each other as some sort of reference to a spatial extension. Spatial entities are placed in a short-term memory, so that consecutive segments that belong to a complex gesture (imagine a person "drawing" the outline of a square by gesture) can be considered.

The object knowledge-base contains representations of virtual objects based on spatial entities. Each object is described by a set of entities (like axes) with spatial relations between them. This representation should be explicit about the qualitative relation of different extensions, indicating, for example, that one axis of an object "dominates" another axis. A further point to be emphasized is the decomposition of objects which should be taken into account by using a structured object representation.

To establish reference between gesture and an object in the virtual world, spatial properties of the gesture have to be matched against object properties with the best-matching object being selected as the referent. This process is called iconic mapping by Sparrell and Koons (1994). Our study shows that iconic mapping should not be quantitative, at least not in a direct way. For example, it should be possible to map a 20cm linear movement onto an axis that is 1m long. The mapping process should not rely on a full match between all the geometric properties of a single object and the form features, because gesture often reflects just an abstracted conception of an object.

6 Conclusion and Further Work

In this paper, we presented a study on the description of virtual objects with iconic gestures. By explicitly analyzing gesture shape and geometric object attributes we pursued the question how iconic gestures are utilized in a Virtual Environment scenario. We found that such gestures convey geometric attributes by abstraction from the complete object shape. Spatial extensions in different dimensions and roundness constitute the dominant "basic" attributes in our corpus. Complex objects were found to be decomposed and described by independent consecutive gestures. We further observed that geometrical attributes can be expressed in several ways using combinations of movement trajectories, hand distances, hand apertures, palm orientations, hand-shapes and index finger direction.

Based on our observations, we proposed a rough model of an iconic reference recognizer that will be our starting point for refinements and a system implementation. Further work needs to be done with respect to the representation of objects using spatial entities and their relations. A prototypical implementation is planned to be embedded in our existing VE framework that already provides mechanisms for scene visualization, sensor data analysis, and feature integration.

Acknowledgement

The authors would like to thank Ian Voss who prepared and carried out the experiments reported here in the context of his master's thesis.

Timo Sowa	*Phone: +49-521-106-2921*	*Email: tsowa@techfak.uni-bielefeld.de*
Ipke Wachsmuth	*Phone: +49-521-106-2924*	*Email: ipke@techfak.uni-bielefeld.de*
Faculty of Technology – AI Group	*Fax: +49-521-106-2962*
University of Bielefeld
33594 Bielefeld, Germany

References

Battison, R. *Lexical borrowing in American Sign Language.* Silver Spring, MD: Linstok Press, 1978.

Beattie, Geoffrey, and Heather Shovelton. “Do Iconic Hand Gestures Really Contribute Anything to the Semantic Information Conveyed by Speech? An Experimental Investigation.” *Semiotica* 123 1/2 (1999): 1-30.

Cohen, Philip R., Michael Johnston, David McGee, Sharon L. Oviatt, James A. Pittman, Ira Smith, Liang Chen, and Josh Clow. “QuickSet: Multimodal Interaction for Distributed Applications.” *Proceedings of the Fifth Annual International Multimodal Conference.* New York: ACM Press, 1997. 31-40.

Efron, David. *Gesture, Race and Culture.* The Hague, Paris: Mouton, 1972. Reprint from *Gesture and Environment.* New York: King’s Crown Press, 1941.

Ekman, Paul, and Wallace V. Friesen. “The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding.” *Semiotica* 1 (1969): 49-98.

Feyereisen, Pierre, and Jacques-Dominique de Lannoy. *Gestures and Speech*. Cambridge, Paris: Cambridge University Press, Editions de la Maison des Sciences de l´Homme, 1991.

Feyereisen, P., M. van de Wiele, and F. Dubois. “The Meaning of Gestures: What Can be Understood Without Speech*?” European Journal of Cognitive Psychology* 8 (1988): 3-25.

Hadar, Uri. “Two Types of Gesture and their Role in Speech Production.” *Journal of Language and Social Psychology* 8 (1989): 221-228.

Hadar, Uri, and Brian Butterworth. “Iconic Gestures, Imagery, and Word Retrieval in Speech.” *Semiotica* 115 1/2 (1997): 147-172.

Kendon, Adam. “Do Gestures Communicate?: A Review.” *Research on Language and Social Interaction* 27 3 (1994): 175-200.

---. “Gesticulation and Speech: Two Aspects of the Process of Utterance.” *The Relationship of Verbal and Nonverbal Communication*. Ed. M.R. Key. The Hague: Mouton Publishers, 1980. 207-227.

---. “Some Relationships Between Body Motion and Speech.” *Studies in Dyadic Communication*. Ed. Aron Wolfe Siegman, and Benjamin Pope. New York: Pergamon Press, 1972. 177-210.

---. “Some Uses of Gesture.” *Perspectives on silence.* Ed. D. Tannen, and M. Saville-Troike. Norwood, NJ: Ablex, 1985.

Kita, Sotaro. “The Temporal Relationship between Gesture and Speech: A Study of Japanese-English Bilinguals.” Master’s Thesis. University of Chicago, 1990.

Kita, Sotaro, Ingeborg van Gijn, and Harry van der Hulst. “Movement Phases in Signs and Co-speech Gestures, and Their Transcription by Human Coders*.” Gesture and Sign Language in Human-Computer Interaction*. Ed. Ipke Wachsmuth, and Martin Fröhlich. Berlin: Springer, 1998. 23-35.

Koons, David B., Carlton J. Sparrell, and Kristinn R. Thórisson. “Integrating Simultaneous Input from Speech, Gaze and Hand Gestures.” *Intelligent Multimedia Interfaces.* Ed. Mark T. Maybury. Cambridge: MIT Press, 1993. 257-276.

Krauss, Robert M., and Uri Hadar. “The Role of Speech Related Arm/Hand Gestures in Word Retrieval.” *Gesture, Speech, and Sign*. Ed. Lynn S. Messing, and Ruth Campbell. Oxford: Oxford University Press, 1999. 93-116.

Latoschik, Marc, Martin Fröhlich, Bernhard Jung, and Ipke Wachsmuth. “Utilize Speech and Gestures to Realize Natural Interaction in a Virtual Environment.” *Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society, Vol. 4.* IEEE, 1998. 2028-2033.

McNeill, David. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: The University of Chicago Press, 1992.

Peirce, Charles Sanders. *Collected Papers of Charles Sanders Peirce*. Ed. Charles Hartshorne, and Paul Weiss. Volumes 1-8. Cambridge: The Belknap Press of Harvard University Press, 1965.

Prillwitz, Siegmund et al. *HamNoSys Version 2.0 - Hamburg Notation System for Sign Languages - An Introductory Guide*. Hamburg: Signum Press, 1989.

Rimé, Bernard, and Loris Schiaratura. “Gesture and Speech*.” Fundamentals of Nonverbal Behavior*. Ed. R.S. Feldman, and R. Rime. New York: Press Syndicate of the University of Cambridge, 1991. 239-281.

Sparrell, Carlton J., and David B. Koons. “Interpretation of Coverbal Depictive Gestures.” *Working Notes of the AAAI Spring Symposium on Intelligent Multi-Media Multi-Modal Systems, Stanford, U.S., 21-23 March 1994.* Stanford University, 1994. 8-12.

Wachsmuth, Ipke, and Yong Cao. “Interactive Graphics Design with Situated Agents.” *Graphics and Robotics.* Ed. Wolfgang Straßer, and Friedrich M. Wahl. Berlin: Springer, 1995. 73-86.

Wachsmuth, Ipke, and Martin Fröhlich, eds. *Gesture and Sign Language in Human-Computer Interaction*. Lecture Notes in Artificial Intelligence, Vol. 1371. Berlin: Springer, 1998.