Univ Access Inf Soc (2015) 14:333–349
DOI 10.1007/s10209-014-0386-8

LONG PAPER

WikiNect: image schemata as a basis of gestural writing for kinetic museum wikis

Alexander Mehler · Andy Lücking · Giuseppe Abrami

Published online: 2 September 2014

Abstract This paper provides a theoretical assessment of gestures in the context of authoring image-related hypertexts by example of the museum information system WikiNect. To this end, a first implementation of gestural writing based on image schemata is provided (Lakoff in Women, fire, and dangerous things: what categories reveal about the mind. University of Chicago Press, Chicago, 1987). Gestural writing is defined as a sort of coding in which propositions are only expressed by means of gestures. In this respect, it is shown that image schemata allow for bridging between natural language predicates and gestural manifestations. Further, it is demonstrated that gestural writing primarily focuses on the perceptual level of image descriptions (Hollink et al. in Int J Hum Comput Stud 61(5):601–626, 2004). By exploring the metaphorical potential of image schemata, it is finally illustrated how to extend the expressiveness of gestural writing in order to reach the conceptual level of image descriptions. In this context, the paper paves the way for implementing museum information systems like WikiNect as systems of kinetic hypertext authoring based on full-fledged gestural writing.

Keywords Gestural writing · Museum information system · Kinetic hypertext authoring · Kinect · Image schemata · Deictic and iconic gestures

A. Mehler (✉) · A. Lücking · G. Abrami
Text Technology Lab, Goethe University Frankfurt, Frankfurt, Germany
e-mail: Mehler@em.uni-frankfurt.de

A. Lücking
e-mail: Luecking@em.uni-frankfurt.de

G. Abrami
e-mail: abrami@em.uni-frankfurt.de

1 Introduction

More and more *human–computer interfaces* (HCI) are designed to incorporate non-verbal communication means [23]. One reason is that tasks are often performed more easily by using, for example, gestures instead of keyboard input or voice control [35]. Another reason is that a growing number of application domains invite users to give feedback (e.g., by filling out questionnaires). A domain that is almost never equipped with multimodal HCIs but requires a great deal of user feedback is that of *museum education*. In this paper, the authors propose gestural writing as a means of controlling information systems for museum education with the aim of getting this feedback.

Generally speaking, both the information seeking and the technical side of HCI are part of the principles and standards of *American museum education* [12, p. 7]:[1] (i) "information gathering and assessment provide evidence of visitor learning and the museum's impact"; (ii) "appropriate technologies are used to expand access to knowledge and self-directed learning." Reference points of assessing the quality of technologies are additionally described by ISO 9241, "namely *efficiency* and *effectiveness* and, moreover, *learnability* and *intuitivity*." (quoted after [55], emphasis in original)

Regarding the aim of (i) efficiently learnable multimodal HCIs and (ii) the field of museum information systems, *WikiNect* is proposed as a system that allows for authoring hypertexts by means of *gestural writing* [38]. The aim of WikiNect is to enable the authoring of museum

[1] For German counterparts of these requirements see for instance [8, p. 8].

wikis by means of gestures. Currently, WikiNect exists in the form of two prototypes based on the Kinect technology.[2] Its usage scenario is given by museums that seek to know what visitors think about their exhibitions. Accordingly, in the context of art exhibitions, the prime domain of WikiNect is the description of images. More specifically, WikiNect aims at enabling visitors to gesturally manifest speech acts whose propositional content informs about the objects (e.g., *paintings*) of an exhibition, their segments, content-related attributes, relations (e.g., *painting A and B share a segment*) and ratings. In this way, WikiNect aims at facilitating on-site feedback about museum exhibitions by means of the so-called non-contact *gestural writing*. Using the MediaWiki technology, WikiNect additionally allows for online collaborative writing regarding the further processing of this feedback. This enables prospective visitors to learn about a museum's exhibitions by gesturally documented experiences of former visitors and their elaborations. To this end, WikiNect integrates three approaches to HCI: the paradigm of *games with a purpose* [54], the principle of *wiki-based collaborative writing* [31] and the concept of *kinetic text-technologies* [38].

This paper deals with the identification of gestures in the context of three constraints to allow for gestural writing in the framework of WikiNect: Firstly, it is required that relevant gestures are instantaneously learnable, so that they reduce the learning effort on the part of museum visitors. Secondly, relevant gestures should be automatically segmentable and identifiable by means of the Kinect technology. According to these two requirements, target gestures should be as simple as possible. Thirdly, relevant gestures should allow for a sort of gestural writing whose expressiveness approximates that of natural language speech acts. According to this requirement, target gestures should be as expressive as possible. In this paper, a set of gestures is specified that aim at fulfilling these conflicting requirements. This is done with the help of the notion of image schemata [29]. Image schemata are basically used to provide expressive iconic gestures as means of predicate selection in gestural writing. They are needed to reduce the search space of gestures that are appropriate for gestural writing in the framework of image descriptions. In this sense, the paper aims at providing a gesture-based HCI that reduces the learning effort of the user while extending the expressiveness of gestural writing. It provides a conceptual framework for designing gesture-based information systems. This is finally needed to pave the way for implementing WikiNect based on full-fledged gestural writing.

As an example, think of a statement like *This painting has the same subject as that painting*. How can one express the propositional content of such a statement by using only gestures? Firstly, the two referents, the paintings, can be identified by pointing gestures. Secondly, in order to express the relation of sharing a subject, one can exploit the imagistic power of image schemata. By evoking, for example, the CONTAINER schema, one can express that the paintings belong to one group (i.e., that they share something). The definition and implementation of this sort of gestural writing together with an assessment of its expressiveness (in this example: *did we really express that both paintings have the same subject?*) is the main contribution of the article.

The paper is organized as follows: First, a short overview of related work is provided (Sect. 2) and briefly the architectural model underlying WikiNect (Sect. 3) is described. A prototypical usage scenario of WikiNect is also provided in order to distinguish five operations of image description: *segmenting*, *relating* (or linking), *configuring*, *attributing* and *rating*. Further, a typology of gestures relevant for HCI is given in Sect. 4. In the main part of the paper, in Sect. 5, a model of gestural writing based on the notion of image schemata is developed and its expressiveness in terms of three levels of image descriptions as distinguished in [20] is rated. Finally, Sect. 6 gives a conclusion and a prospect of future work.

2 Related work

This paper, which is in some parts an extended version of [38], is connected to three areas of related work: hypertext authoring, Kinect-based interfaces and research on gestures. The overview of related work primarily focuses on the first two parts of this triad.

To the authors' knowledge, there is no prior work that targets the authoring of hypertext by means of gestures. Nevertheless, WikiNect consorts with a couple of proposals from the hypertext literature. For example, the learning of creative gestures could provide a means of customization that is needed for adaptive hypermedia [16]. With WikiNect, personalized recommendations based on rating predictions [51] become possible even in the context of art exhibitions. Take, for instance, the learning of user characteristics from social tagging behavior [49]. WikiNect shares with the *Spatial Hypertext Wiki* (ShyWiki) [52] that it allows users not only to organize relations collaboratively among wiki pages, but also to visualize them. Since gestures are visuo-spatial in nature, WikiNect also contributes to setting up a (more apt) multimodal vocabulary for the problem of finding visual notions for spatial hypertext [5].

[2] Both of these prototypes have been developed in the framework of the Text Technology lab at Goethe University [1, 21].

A second area of relevant work relates to the fields of gesture-based navigation, instruction games and exergames. Applications in this area mostly focus on controlling interfaces [11, 42, 46] and supporting users [34, 42], especially those with physical impairments [17, 40, 47]. Common to these approaches is that they use Kinect's 3D depth camera for enabling gesture-based interactions. Cochran [11], for example, describes a Kinect-based user interface for the interaction with a semi-spherical dome with built-in LED lights. To change the color of the lights, users have to produce pointing gestures. Underlying interaction techniques are described in [9]. The added value of non-contact interactions is demonstrated in [42], where a smart kitchen equipped with a music player and a recipe navigator is introduced. In this application, one can control the radio or look up the recipe in a contact-free manner. Another example is described in [46], where methods for scaling, moving and grabbing virtual objects are tested in the scenario of assembling technical constructs by means of moving virtual robots. Finally, user-assisting exergames are described in [47]. This includes a yoga instructor that exercises blind or visually impaired people, where a skeleton tracking algorithm is used to analyze body positions. A related example is the *Super Mirror* of [34], which provides an interface for ballet dancers also by means of skeleton tracking. Another example is EMERGANZA, a prototype in the area of Life and Medical Sciences Health [2]. EMERGANZA is used in the RIMSI[3] project "for the simulation and training of medical and paramedical personnel in emergency medicine" [2]. It can be controlled by Kinect and was designed as a free-roaming game. By detecting the configuration of the hand with a skeletal tracking system, the user can point at objects in the 3D environment.

The relation between gestures and cognitive structures has been a topic in various works. Most prominently, in speech-and-gesture production research, it is a wide-held view that gestures manifest mental representations. This view is captured in the metaphorical phrase of de Ruiter [13], namely that gestures are "Postcards from the mind." Adhering to metaphors, image schemata and the corresponding conceptual metaphor theory have been employed in describing and analyzing metaphorical gestures (or non-metaphorical gestures that co-occur with metaphorical speech)—see, for instance, [39] and some of the works cited there. Cognitive structures have also been used in non-metaphorical semantic accounts for co-verbal iconic gestures. In [32], for instance, iconic gestures are analyzed as exemplificators of intensional perceptual structures.

[3] www.micc.unifi.it/rimsi.

3 The WikiNect architectural model

In this section, the architecture of WikiNect (see Fig. 1) is briefly described together with a prototypical usage scenario of using it. Generally speaking, WikiNect has two access points. One is *on-site* in the museum, the other is *online* via the web—the user interface and the web front-end, respectively (see Fig. 1). Each of these access points refers to the same database using a MediaWiki as its web front-end. The on-site authoring interface is controlled by the session management module, which provides image description templates (focusing on segmenting, linking, configuring, attributing and rating images). Museum visitors can author WikiNect entries in a WikiNect-session. Part and parcel of such a session is that a user can operate the system by gestures as will be defined and exemplified below. A session defines a subject (a single image, a group of images or a complete workplace of images defined by the user) together with a task in image description (segmenting, rating, etc.). Since image descriptions cannot yet be fully expressed by means of gestures (as will be shown below), WikiNect contains a language model based on speech recognition together with a virtual keyboard. The keyboard can be controlled by means of pointing gestures or by means of the sign alphabet to enter free text. The on-site produced hypertext entries (image segments, their relations and descriptions) can also be accessed online through the web front-end. Since this front-end is detached from on-site sessions, users can further process the content of WikiNect by analogy to Wikipedia. In this way, WikiNect enables users to provide image descriptions with respect to museum exhibitions in a collaborative manner, while the on-site interface attracts users to take part in WikiNect and to continue their image descriptions. Last but not least, the contactless interface in terms of gestural writing is needed to guarantee a low entry point together with a low learning effort. For more details on WikiNect's architecture, see [38].

A typical application scenario of WikiNect is sketched in Figs. 2, 3, 4 and 5.[4] An exhibition visitor (the user) is enabled to provide information about a selection if images he/she has seen as part of the exhibition. Since his/her movements are tracked by a Kinect controller, information can be given in a gestural way. For instance, the user might find that the two outer portraits look similar. He/she can express a respective proposition by selecting both images by means of *deictic gestures* and *linking* them in terms of an *iconic gesture* that pictorially depicts the similarity relation. Figure 2 displays an example for such a gestural linking sequence: two pointing gestures single out the

[4] The image files used in the figures are taken from http://commons.wikimedia.org and belong to the public domain.

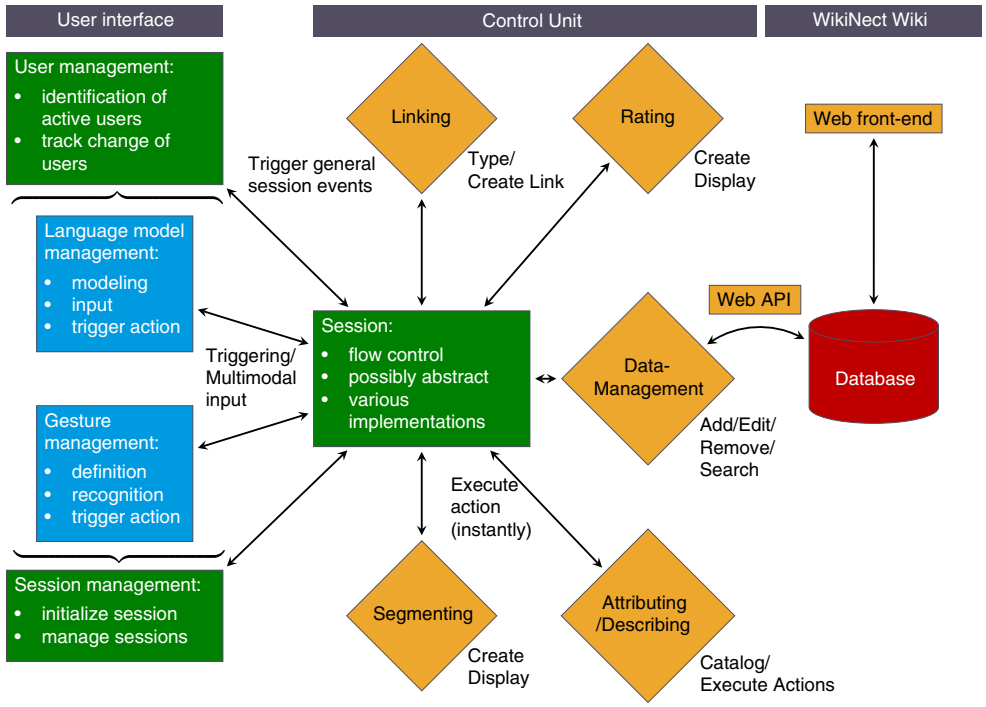

Fig. 1 The architecture model of WikiNect

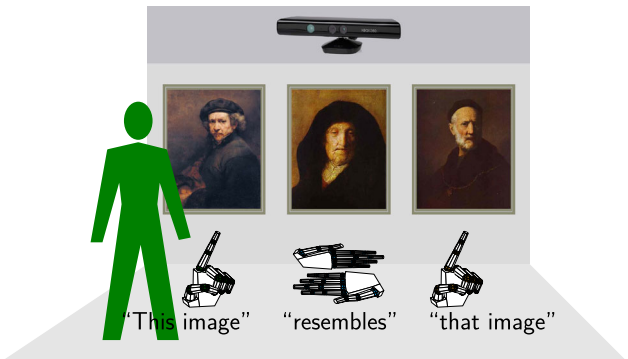

Fig. 2 Linking

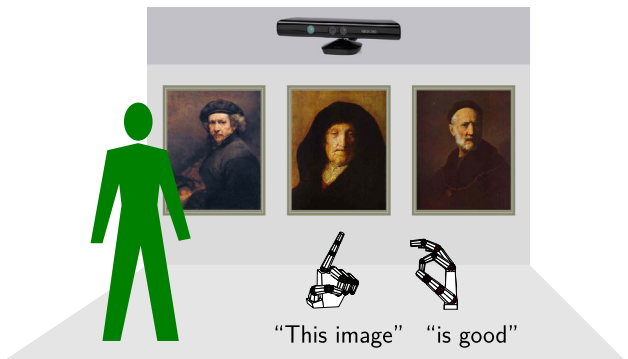

Fig. 3 Rating

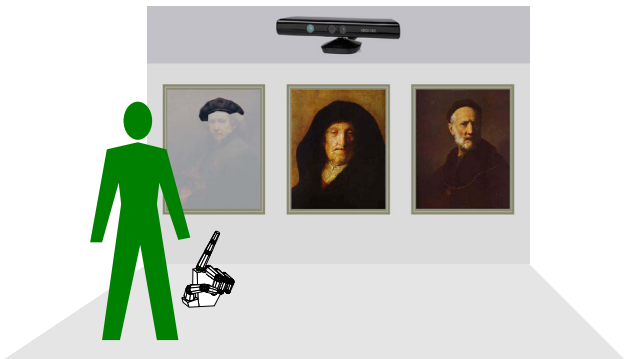

Fig. 4 Segmenting

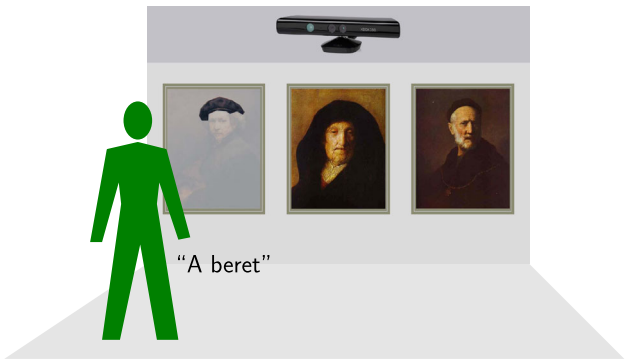

Fig. 5 Attributing

portraits under discussion, which in turn are related by a gesture mimicking the equals sign.

In order to express that an image is particularly good, the user can draw on conventionalized or *symbolic* gestures like *Thumbs-Up* or the "OK" sign—see Fig. 3 for such a *rating* example.

Furthermore, a gesture can also be used for *segmenting* images. For instance, the head covering of the leftmost portrait in Fig. 4 can be cut out by a gesture used as a cutting tool. Once a segment of an image has been cut out, the user might want to label it. In case of a predicate like "beret," as shown in Fig. 5, it is unclear, however, how this *attribution* can be made purely in terms of gestures. This is part of the problems discussed in this paper. Against the backdrop of deictic, iconic and symbolic gestures, it is assessed by what means and to what extent propositions for linking, rating, segmenting, configuring and attributing images can be expressed purely gesturally. To this end, in the subsequent Sect. 4, the gesture backdrop is elaborated in more detail.

4 Classification of gestures in HCI

In order to get an overview of the kinds of gestures that are relevant for HCI and especially for WikiNect, a "typology of gestures" [44] is provided. The typology starts with distinguishing gestures as signs from gestures as actions. It then distinguishes gestural signs according to whether they are codified or not. Taking up the usage examples from the preceding section (cf. Figs. 2, 3, 4, 5), the "OK" gesture is a codified one, while iconic or deictic gestures are rather spontaneous or creative. Since gestures that are part of HCI differ in certain respects from gestures that occur as part of natural language communication, the typology should make this difference explicit. As a result, eight classes of gestures are identified, out of which five are used in HCI contexts.

First of all, the term 'gesture' is understood as to denote hand-and-arm movements. In this understanding of gesture, facial expressions or body postures are excluded. However, this notion includes hand-and-arm movements that are *actions* or gestures of a *sign language*.[5] The term 'action' refers to hand-and-arm movements that involve the use of a concrete object, be it intransitively or transitively (that is, applied to a second object). For instance, the action of cutting bread (involving the objects *knife* and *bread*) is distinguished from a gesture simulating the cutting of bread (involving no objects at all). In HCI, actions prevail in terms of manipulations of some interface entity (say, a scroll bar or button). For this reason, they have been dubbed *manipulators* [45]. They owe their name to interfaces which provides entities that can be manipulated by users—say, scroll bars, clickable icons, movable objects, and the like. The great advantage of manipulators for HCI is that they provide direct feedback [7].

The second important kind of gestures are *codified gestures*. A gesture is codified if its form-meaning mapping is regimented by a symbolic convention. With regard to HCI applications, the most important class of codified gestures is the class of *semaphores*. Semaphoric gestures are HCI gestures that make up a predefined set of stylized gestures [45, p. 173]. Many touch gestures used to operate touchpads or touch screens belong in this category.[6] They become stylized due to their widespread usage in the increasing field of personal technological devices. Emblematic gestures, or simply emblems, can be regarded as maximally stylized, since their form-meaning relation is assumed to be lexicalized, though culture-specific (a standard example being the victory sign or the "OK" gesture from above).

Codified gestures are contrasted to non-codified, non-stylized, or simply *creative* gestures. The class of creative gestures can be further distinguished according to their function in discourse (cf. gesticulations in the sense of Kendon [26] and McNeill [37]). The resulting typology is given in Fig. 6, where the kinds of gestures that according to Karam and Schraefel [25] predominate in HCI are indicated by light-squared boxes.

Following mainly McNeill [37], creative gestures—what he calls *gesticulations*—can be partitioned into beats, iconic gestures, metaphoric gestures, and deictic gestures. Deictics are pointings, either concrete—that is, to something in the perceptible environment, like pointing at images as illustrated, e.g., in Fig. 3 above—or abstract—that is, to some "semantically loaded" part in gesture space. Iconics are said to resemble, mimic or simulate their referent, while metaphorics provide a spatio-visual depiction of something abstract. The depiction of the similarity relation from the application example in Fig. 2 is performed by a gesture of this kind. Finally, beats are rhythmic movements that emphasize accompanying verbal units or indicate a structuring of the co-verbal utterance (like in enumerations). See Ekman and Friesen [14] for a slightly different typology, mainly with regard to beats, and see Müller [41], Streeck [53] and Lausberg and Sloetjes [30] for elaborating, *inter alia*, on iconic gestures (Figs. 7, 8, 9, 10).

[5] In English, there is no terminological distinction between everyday gestures and the manual signs of a sign language, which makes it a bit more inconvenient to hold them apart. In German, for example, a distinction is made between *Geste* (gesture) and *Gebärde* (sign language sign).

[6] For an example, see the track pad gestures from www.apple.com/magictrackpad, accessed October 17, 2013.

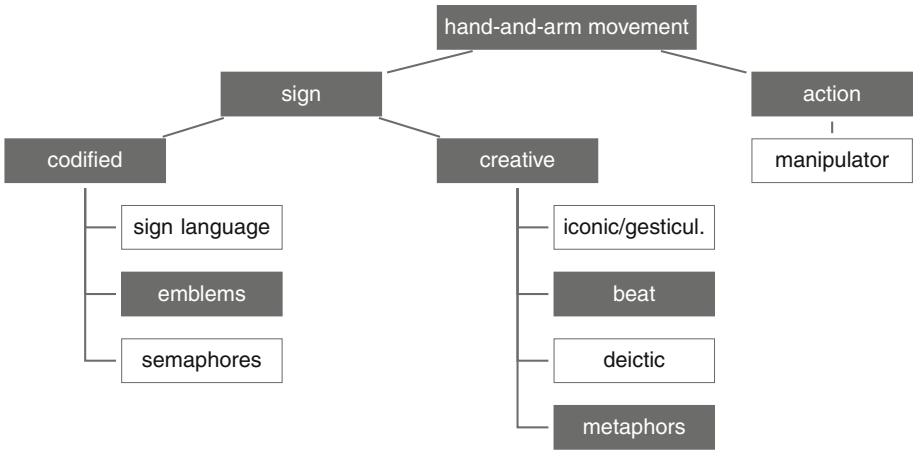

Fig. 6 A typology of gestures: the light-square entries are kinds of gestures that are used in HCI contexts [25]

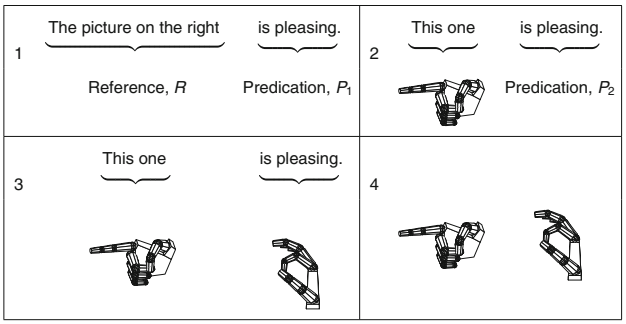

Fig. 7 From utterances via multimodal behavior to gestural writing in four steps (see [38])

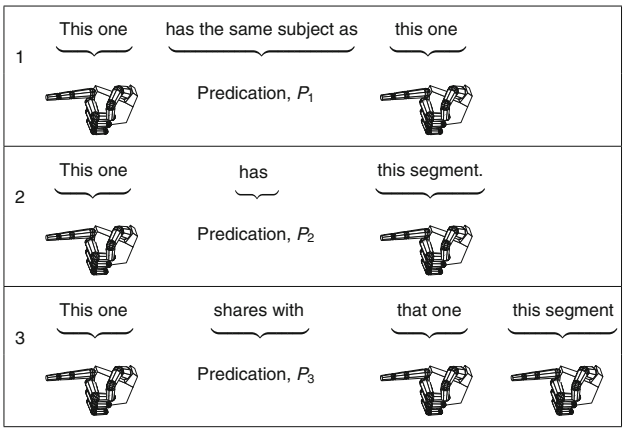

Fig. 8 Transitivity patterns of speech–gesture interaction (see [38])

The typology of gestures described so far is the starting point of introducing gestural writing in the next section. As will be shown, deictic and iconic gestures are the basic means of this kinetic text-technology. The relevant gestures are depicted in Figs. 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31. In these

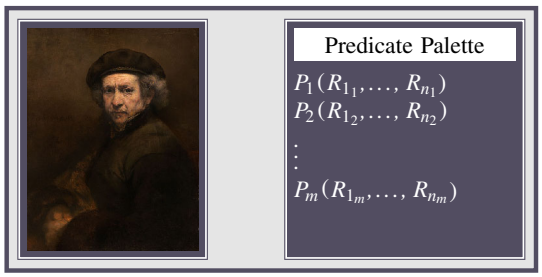

Fig. 9 Workspace with a palette of predicates (image source: www.commons.wikimedia.org)

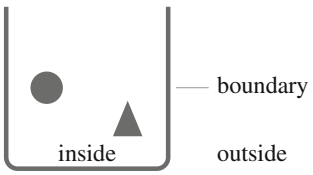

Fig. 10 The *container* schema

figures, left and right hands are distinguished by different colors. If just one hand is used, temporal dynamics is indicated by shaded hands that mark previous positions of the hand. Hands are either drawn from the perspective of the gesturer or from the perspective of an observer standing opposite.[7]

5 Gestural writing

In [38] it is argued that gestures can become associated with referential expressions or verbal predicates to form the

[7] The hands are drawn by means of the *Sketch* system (www.frontiernet.net/~eugene.ressler/sketch.html) and built on the hand model of Eugene Ressler.

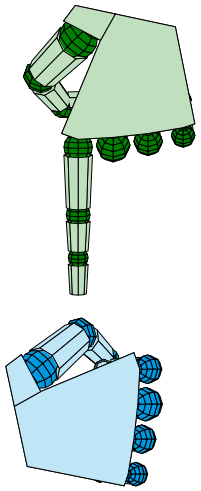

Fig. 11 Container gesture

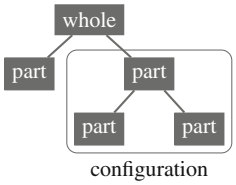

Fig. 12 The *part–whole* schema

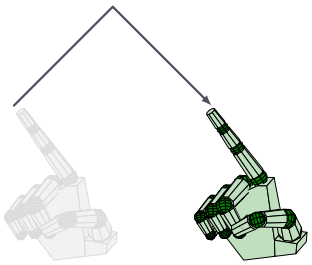

Fig. 13 Part–whole gesture

Fig. 14 The *link* schema

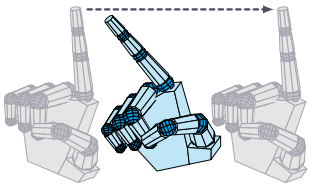

Fig. 15 Link gesture

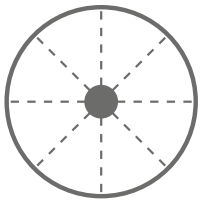

Fig. 16 The *center–periphery* schema

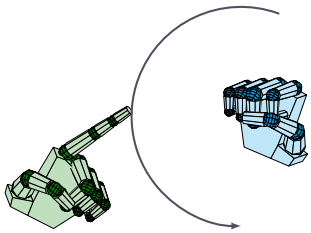

Fig. 17 Center–periphery gesture

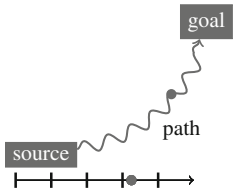

Fig. 18 The *source–path–goal* schema

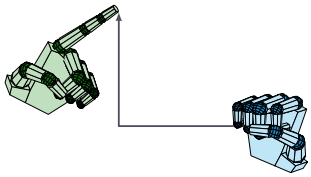

Fig. 19 Source–path–goal gesture

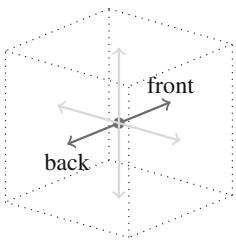

Fig. 20 The *front–back* schema

so-called speech-and-gesture ensembles (see also [27] and [33]). Generally speaking, a speech-and-gesture ensemble is a multimodal supersign that consists of at least one verbal unit and one gesture that tend to co-occur in multimodal communication. The association of gestural and verbal signs paves the way to monomodal gestural communication, where the latter are partially replaced by the former. From this point of view, one can think of a scale that is spanned by two extremal cases of exclusively verbal input on the one hand and exclusively gestural input on the other. The transition between these two endpoints is

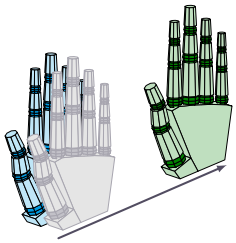

Fig. 21 Example: forward gesture

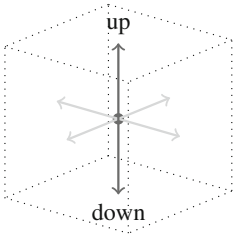

Fig. 22 The *up–down* schema

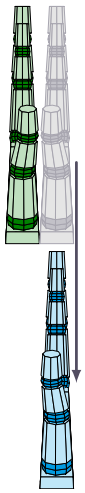

Fig. 23 Example: downwards gesture

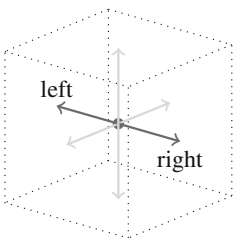

Fig. 24 The *left–right* schema, 2

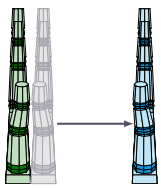

Fig. 25 Example: rightwards gesture

illustrated in Fig. 7. It starts from a speech act whose proposition reads as a statement that a referent x has some property P (Row 1). Both parts of the proposition, reference and predicate, can be manifested by a multimodal utterance,

Fig. 26 The *contact* schema

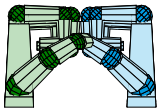

Fig. 27 Contact gesture

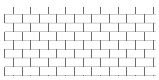

Fig. 28 The *texture* schema

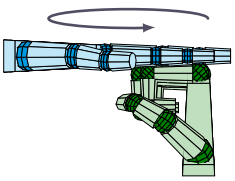

Fig. 29 Texture gesture

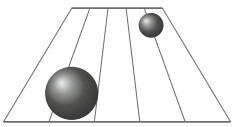

Fig. 30 The *near–far* schema

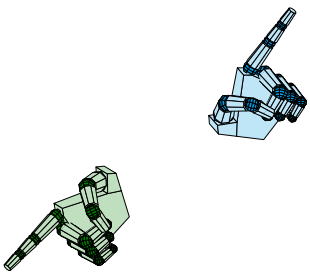

Fig. 31 Near–far gesture

where the shortened verbal statement is accompanied by gestures (Rows 2 and 3). If gestures are associated with verbal correlates, it becomes possible to omit the verbal statement and to express the proposition by purely gestural means (Row 4). In this example, a deictic gesture is connected to the reference part of the proposition, while an emblematic gesture is connected to its predication. The bridging between verbal and gestural input is provided by co-verbal gestures as

constituents of speech-and-gesture ensembles. The authors refer to these bridging gestures as *transient gestures*, or simply *t-gestures* [38]. In what follows, the term *gestural writing* is used when dealing with writing documents (e.g., articles in a wiki) exclusively by means of gestures. It is hypothesized that *t*-gestures are candidates for realizing this sort of writing.

More specifically, *t*-gestures are seen as candidate elements of a repertoire of the so-called *gestograms*. By analogy to pictograms, gestograms are defined to be iconic or indexical gestures that have a referential or predicative meaning. The idea is that gestograms are *t*-gestures that are easily communicated without verbal affiliates. This notion is connected to a novel task in machine learning, which aims at identifying gestograms as constituents of speech-and-gesture ensembles in multimodal data streams. In what follows, a set of gestures is analyzed as candidate gestograms. In order to stress the central role of gestograms in gestural writing, the requirements analysis starts with the following statement:

Requirement 1 In gestural writing proper, indexical or iconic gestograms are the only means to gesture parts of a propositional act.

In what follows, gestural writing is specified based on gestograms in terms of *image schemata* [10, 29]. Image schemata are the means to bridge between the semantics of natural language units on the one hand and gestograms on the other. Further, the notion of a propositional act is utilized as part of speech acts [50] to realize gestural writing by means of WikiNect. It is shown that this kind of writing requires a transformation of a symbolic code (as exemplified by natural languages) into a combined indexical and iconic code with symbolic add-ons. The main implication of the analysis is that gestural writing primarily focuses on the perceptual level [20] of image description. A second finding is that image schemata allow for bridging between natural language predicates and gesto-grammatical manifestations. A third implication is that by exploring the potential of image schemata with respect to metaphor formation, one can go beyond this limit, but to the prize of symbolic codings of predicate selections. In a nutshell: It is argued that gestural writing comes into reach with the development of non-contact interfaces subject to a trade-off between the expressiveness of gestograms (compared to the one of natural languages) and the effort of learning them.

Note that the status of image schemata in the presented approach is questionable for several reasons. In the literature, it is argued that the derivation of image schemata rests on a circular argumentation [19, 36] (quoted after [28]), that the notion of embodiment, which is part and parcel of the definition of image schemata, is overstated [18] and that it is unclear how to distinguish image schemata from "usual" cognitive structures. The latter identification problem is more evident if one tries to unify the proposed image schemata of, (e.g., [10, 24, 29]). Moreover, from the point of view of gesture modeling, one may argue against the notion of metaphoric gestures [6], which in the present approach play a central role in extending the expressiveness of gestural writing. In contrast to these arguments and what regards the cognitive plausibility of image schemata, Rohrer [48] summarizes evidence on the neural basis of image schemata.

In any event, it is important to emphasize that gestural writing is not inevitably tied to image schemata. In principle, it can be based on many cognitive theories. However, since the theory of image schemata provides an intuitively accessible representation format, the discussion is framed in terms of such schemata.

Generally speaking, repertoires of gestograms have to be distinguished from letter substituting codes as exemplified by fingerspelling. Obviously, fingerspelling does not provide candidate codes for implementing gestural writing. One reason is that it has to be learned as an alphabetical writing system. For the same reason, "writing into the air" and contact-free manipulations of virtual keyboards are not instances of gestural writing. Rather, *t*-gestures are sought that, due to their potential to be associated with verbal units in a usage-based manner, can be used as *gestograms* and therefore allow for circumventing writing according to an alphabetical code.

To introduce the notion of gestograms, first an analysis of propositions is provided. The starting point is given by sentence-related speech acts $F(RP)$ as analyzed by Searle [50], where F indicates the illocutionary force of the act, while R indicates the reference and P the predication of the underlying proposition. It is assumed that the illocutionary role of a proposition in HCI is determined by the specific task the system helps to accomplish. That is, in HCI propositional contributions from the user are typically embedded under a *command* or *request* illocution. Up to that point, it is assumed that F indicates an assertion. According to Searle [50], it is assumed that R identifies a single entity (and not, for example, a group of entities), while P expresses a single predicate, which is attributed to the entity identified by means of R and cannot be decomposed into a sequence of other predicates. In this scenario, which is exemplified in Fig. 7, gestural writing has to provide gestures for filling in the variables R and P.

Evidently, propositional acts as manifested by natural languages are far more complex than the ones in Fig. 7. This is demonstrated in Fig. 8, in which Row 1 exemplifies a speech act that contains two references linked by a transitive predicate expressing a symmetric relation. In contrast to this, Row 2 exemplifies an asymmetric relation, while in Row 3, three references are linked by a ditransitive predicate. Obviously, one can think of any complex

predicate P that, though being manifested by a single sentence, can achieve the complexity of a predicate manifested by a natural language discourse. In order to circumvent the corresponding variety of complex predicates, it is assumed that propositions to be manifested by means of WikiNect are realized by sentences that contain exactly one (non-complex) predicate.

To get a more systematic account of gestural manifestations of propositional acts, the schema

$$P(R_1, \ldots, R_n) \quad (1)$$

is used for which it is assumed that P is a predicate of arity n such that the position of its ith argument—as identified by R_i, $1 \leq i \leq n$,—is bijectively mapped onto a corresponding *thematic role* (*deep case* [15] or *relational category* [29, p. 285] as exemplified by *agent* and *instrument*).[8] According to this scheme, two functions of gestures in WikiNect can be distinguished:

1. *Gestures for manifesting speaker connections* (in the sense of Barwise and Perry [3]):
   (a) Firstly, *predicational gestures* are needed for manifesting the *predicate P*.
   (b) Secondly, *deictic gestures* are needed for identifying P's arguments.

Henceforth, one can speak of *Model (1)* of gestural writing when dealing with implementations of this sort of gestural writing. It requires to select a predicate P in conjunction with a preconfigured set of arguments, which have to be identified step by step. The latter preconfigured set of arguments will be called *predicate palette* throughout this paper (see Fig. 9). Model (1) is problematic in two respects:

– *Problem 1* The selection of P is both *indexical* (in that the writer needs to point at the predicate to be selected as part of the predefined palette $\mathscr{P}$ of predicates) and *symbolic* (in that the writer needs to read the symbolic presentation of P as part of $\mathscr{P}$ before she can select it). That is, Model (1), as depicted in Fig. 9, combines an indexical, gestural code with a symbolic, alphabetical code. In this way, Model (1) departs from the requirements analysis that asks for *gestograms* as means to gesture any part of a propositional act. Below, Problem 1 is addressed by means of introducing iconic gestures for the selection of predicates.
– *Problem 2* Model (1) requires to strictly follow the configuration of thematic roles as established by the predicate P. This induces a sort of inflexible, unnatural writing that the authors also aim at overcoming below.

As long as gestural writing is based on Model (1), the thematic roles of P's arguments can be determined by the linear order of corresponding reference acts. Things are more complex if this order is not followed by the gestural linearization of the underlying speech act. This is exemplified by topicalization that varies the order of arguments depending on what is topicalized. Generally speaking, if the order of references is varied, it is necessary to distinguish two additional functions of gestures in gestural writing:

2. *Gestures for manifesting propositional configurations of arguments*:[9]
   (a) If the order of the references $R_{i_1}, \ldots, R_{i_n}$ does not code their thematic role, WikiNect has to provide *configuring gestures* to manifest the configuration of P_i's arguments. By analogy to drawing by numbers, this may be done by a mapping of the i_jth reference to its corresponding argument position.
   (b) Alternatively, *role assignment gestures* may be introduced that map arguments of predicates onto their thematic roles by whatever temporal order.

With the help of configuring gestures (2a), the role of an argument is specified by its mapping to an argument slot of Schema 1 that is uniquely connected to a certain role. Using role assignment gestures (2b), this assignment is done without assuming any linear order of the arguments of P. Any such ordering (according to 2a or 2b) is indispensable if P expresses an asymmetric relation among at least two arguments. Note that it is not assumed that thematic roles occur at most once in a sentence. In cases in which the same role occurs repeatedly, role assignment gestures need to be accompanied by configuring gestures that disambiguate them.[10] Henceforth, one may speak of *Model (2)* that makes use of variant (2a) or (2b) in addition to Model (1).

So far, two candidates of gestural writing have been described: Model (1) and Model (2). Note that unlike the former, the latter overcomes Problem 2 in that it allows for any order of identifying the arguments of a predicate. However, Model (2) still faces Problem (1) since it does not say anything about selecting predicates other than done before, that is, by combining indexical gestures (pointing at

[8] Evidently, Schema 1 is oversimplifying compared to the complexity of natural language propositions. However, to get started, this analogy to Searle's analysis of propositions is utilized.

[9] In the sense of the PART–WHOLE schema of cognitive semantics—see Lakoff [29].

[10] An example is given by elliptic sentences like "The postman was first attacked with a knife and then with a scissors."

P) and symbolic text processing (in terms of identifying *P* as part of the predicate palette—see Fig. 9). Further, even if one concentrates on sentence-related speech acts and reduces the set of selectable predicates according to a subset of frequent ones, Model (2) will always fail to achieve the expressiveness of a natural language. To see this, look at three candidate instantiations of *P* taken from the image description of Rembrandt's self-portrait of 1,659 as published on Wikipedia[11] (see Fig. 9)—subordinated examples paraphrase parts of their super-level counterparts:

1. "Rembrandt is seated [...]."
2. "The most luminous area [...] is framed by a large beret [...]."
   (a) *This area is luminous.*
   (b) *This area is more luminous than any other area [of the picture].*
   (c) *This area is below that area.*
   (d) *That area shows a beret.*
   (e) *The beret is large.*
3. "The picture is painted in a restrained range of browns and grays [...]."
   (a) *This area is painted in browns.*
   (b) ...
   (c) *That area is painted in grays.*
   (d) *The browns and grays [of the areas] form a range.*
   (e) *The range is restrained.*

In terms of gestural writing, Example (1) requires a segmentation of Rembrandt's contour. This segmentation manifests a reference act by drawing a corresponding polygonal line by means of indexical gestures. Example (1) additionally requires an iconic gesture in order to predicate that the person identified by the latter segmentation is seating. Examples of this sort can be straightforwardly reconstructed in terms of gestural writing. The reason is that sitting is a concrete, spatially organized action that is salient in human experience, an action which, therefore, can be gestured iconically—possibly in conjunction with body movements. This analysis does not hold for Example (2). In this case, it is assumed that the alphabetically ordered sequence of Examples (2-a)–(2-e) is entailed by Example (2). From this point of view, the four predicates of this sequence can be considered stepwise. Starting with the predication (2-a), it is noted that as a quality, *luminosity* is an instance of Peirce's [43] category of firstness and as such a concept that is not spatially organized. Thus, by analogy to abstract terms, this concept cannot be gestured iconically. In order to stress the significance of this finding, it is reformulated in terms of a statement of the requirements analysis of gestural writing:

Requirement 2 Gestural writing in the narrow sense requires that a predicate to be manifested denotes a spatial or temporal property or relation.

Looking at statement (2-b), which contains a universal quantifier, one encounters the next challenge, since there is no single iconic gesture to denote the universe over which is quantified for the given situation. Obviously, it is very laborious to segment each relevant area of the picture and to relate it to the most luminous area in terms of the relation *is less luminous than*. This leads to the next requirement:

Requirement 3 In order to keep gestural writing simple, a proposition should not contain universal quantifiers.

Statement (2-c) is unproblematic: starting from the notion of an image schema [29, Chap. 17], it reminds of the so-called UP–DOWN schema (as depicted in Fig. 22). Figure 23 proposes an iconic gesture as a means to identify this schema. Note that gesturing (2-c) additionally requires two deictic gestures whose referents are linked asymmetrically. Thus, according to the analysis above, instantiating the UP–DOWN schema by example of statement (2-c) presupposes two preceding reference acts of segmenting and pointing or re-pointing already segmented areas as arguments of the schema. It also requires a configuring gesture or a role assignment gesture if the order of the referents is not coded by the temporal order of the reference acts. Once more, an image schema can be utilized to manifest the configuration, now with the help of the SOURCE–PATH–GOAL schema (as depicted in Fig. 18). This requires that the order of arguments is fixed by mapping them onto the path, for example, into the direction of the goal. In this way, each waypoint along the path represents an argument of the predicate. Note that unlike Model (1) and (2), where predicate selection is manifested by a pointing gesture that refers to a symbolic code (i.e., the predicate palette in Fig. 9), the selection is once more made by an iconic gesture as exemplified in Fig. 19, which identifies the SOURCE–PATH–GOAL schema. After having selected the latter schema by the corresponding gesture, the user needs to perform it by drawing a path from argument to argument in the desired order.

To recapitulate the "gesturization" of statement (2-c) as proposed so far:

(a) the user starts with selecting the UP–DOWN schema by means of an iconic gesture (because of its function, this gesture is predicational);
(b) then, two deictic gestures are performed to identify two areas of Rembrandt's self-portrait;
(c) next, the iconic SOURCE–PATH–GOAL schema is identified by a corresponding iconic gesture to indicate that the configuration of the arguments occurs next;

[11] http://en.wikipedia.org/wiki/Self_Portrait_with_Beret_and_Turned-Up_Collar, download: October 21, 2013.

(d) finally, the latter schema is performed by drawing a path from the first to the second argument. In this way, a configurational gesture is performed.

This procedure complies with the requirements analysis of gestural writing given so far since it solely relies on deictic and iconic gestures—without relying on any preconfigured palette of predicates. Moreover, this procedure can be simplified as follows: since WikiNect knows the spatial relation of the segments of a picture, it suffices to select the UP–DOWN schema—afterward, any pair of areas identified deictically can automatically be ordered in terms of up, down, left and right relations. A sole exception is the *front–back* schema because of the two dimensionality of pictures that does not provide the latter kind of prior knowledge.

More challenging than statement (2-c) is the statement (2-d), which can be conceptualized as an attribution of the predicate *beret* to the upper area as identified by the instance of the UP–DOWN schema: Though *beret* is a spatially organized entity, one can hardly assume that it is strongly associated with a *t*-gesture that allows for unambiguously substituting the verbal manifestation of the predicate. The same analysis holds if in (2-d) *beret* is replaced by nouns such as *cap*, *hat*, and *head covering*. What makes examples like these challenging is the openness of the universe from which these nouns can be selected and their seemingly loose association with *t*-gestures. This observation leads to a fourth requirement:

Requirement 4 In gestural writing, predicates have to be identified by means of gestograms or, more specifically, by iconic gestures.

An obvious consequence of Requirement 4 is that gestural writing is seriously limited by the number of predicates it can deal with in a given session, since any predicate to be additionally included induces a corresponding learning effort by associating the predicate with its gestural affiliate. This also means that any predicate that is not covered by the limited set of iconically selectable predicates must be selected in another way (e.g., by means of a predicate palette or by contact-free means of alphabetical writing). It will not belong to gestural writing in the narrow sense of this notion. From this point of view, statement (2-d) is not yet covered by gestural writing, since it requires an iconic-gestural manifestation of the predicate (*is a*) *beret*. A similar analysis holds for Example (3). The reason is that any of the statements (3-a)–(3-e) contains a predicate that is negatively affected by Requirement 2.

5.1 Gestural writing by means of image schemata

According to the requirement analysis given so far, gestural writing is limited to a range of spatio-temporal predicates that are iconically gesturable. In what follows, a subset of such predicates together with their gestural manifestations is proposed. As a corollary, the subset of propositional acts of image description is specified that can be manifested by means of this set of predicates. This is made with the help of the classification of image description operations given by Hollink et al. [20]. Last but not least, it is shown how to extend the expressiveness of gestural writing in terms of the theory of metaphor based as on image schemata.

Section (3) exemplified four operations of image description: segmenting, linking, attributing and rating (segments of) images. Table 1 maps these and related operations onto image schemata as described in cognitive science [10, 29]. Beyond segmentations (of units that denote image elements in the sense of Hollink et al. [20]), configurations of left–right, up–down, front–back, near–far and of contact are distinguished among segments. Further, temporal configurations are presented (as exemplified by statements like "The scene in segment *x* happens before the scene in segment *y*"). In this way, seven operations of configuring subimages are distinguished. Furthermore, the orientation of links (that describe relations in the sense of Hollink et al. [20]) is presented where directed links are identified by means of the SOURCE–PATH–GOAL schema. As this schema is also used for temporal configurations, a differentiating gesture (denoted by the variable ρ) is applied that—by analogy to *radicals* in writing systems—serves for the semantic specification of the otherwise ambiguous gestogram. Henceforth, one may speak of radicals in the case of gestures that serve for specifying the semantics of the corresponding predicational gesture, which selects an ambiguous image schema.

Next, two sorts of rating are allowed for: absolute ratings map single segments, configurations or links onto a predefined scale, while relative ratings relate different objects in terms of being *up* or *down*. Since it is assumed that ratings operate on segments, configurations (i.e., compositions in the sense of Hollink et al. [20]), links, attributions or even on ratings, a further radical to account for this semantic differentiation is needed. Thus, for example, in the case of relative ratings, two consecutive radicals ρ_1 and ρ_2 are needed: By convention, the first determines the function of the CONTAINER schema (e.g., rating) and the second the type of description object (i.e., a segment, configuration, link, attribution or rating) on which it operates. In order to implement absolute ratings, a range of visually depicted, vertically ordered containers can be offered to the user as soon as he/she has decided for this operation of image description. By their order, the containers denote the corresponding rating so that one needs to drag and drop the focal object over the target container in order to rate it. One may speak of *Model 3* of gestural

writing when dealing with implementations of the operations 1–12 of Table 1.

Note that Model 3 allows for the recursive application of image schema-based operations: Rembrandt's portrait (see Fig. 9), for example, can be recursively segmented such that the image element showing Rembrandt's eyes are segmented as subimages of the subimage showing his face (PART–WHOLE schema). The left–right order of the subimages can be explicitly stated by means of a LEFT–RIGHT schema. Using the FRONT–BACK schema, one can state that the folded hands (CONTACT schema) are located in front of the body. The UP–DOWN schema enables stating that the beret is on the head. Further one can state that the center of the picture is occupied by the portrayed person (CENTER–PERIPHERY schema) or that—according to http://en.wikipedia.org/wiki/Self-Portrait_with_Beret_and_Turned-Up_Collar—Rembrandt's portrait alludes to the style of the *Portrait of Baldassare Castiglione* by Raphael (LINK schema). The latter example shows that image schemata can operate on different images (given that they are represented in WikiNect). In this way, it is possible to distinguish between intra- and intermedial relations (see Table 1, Column 7).

A note on the role of image schemata in Model 3: Image schemata are experiential structures that are both recurrently grounded in the spatio-temporal experience and, thus, highly reflected in our language [29]. From the authors' point of view, image schemata relate to Requirement 2, which aims at predicates that denote spatial or temporal relations. In this sense, image schemata provide a cognitive resource of selecting core predicates of gestural writing. Because of being grounded in our spatio-temporal experiences, image schemata allow for being iconically gestured [39] and impose little learning effort. In a nutshell, image schemata can be used to bridge between a core set of predicates that are highly reflected in natural language and their iconic manifestation in gestural writing.

Table 1 distinguishes between attributions of segments, configurations, links, ratings and of attributions

Table 1 Mapping operations of image description onto image schemata in gestural writing: Predicate selection and argument identification occur by means of iconic and deictic gestures, respectively

No.	Image operation	Image schema	Predicate selection [iconic (emblematic)]	Argument identification (indexical)	Argument schema	Intermedial
1.	Segmenting	PART–WHOLE	Gesture 13	(Deictic, seq. of deictic gestures)	$(1, \infty)$	No
2.	Configuring, left–right	LEFT–RIGHT	Gesture 17	(Deictic, emblem, deictic)	(∞, ∞)	Yes
3.	Configuring, up–down	UP–DOWN	Gesture 23 $\circ$ ρ	(Deictic, emblem, deictic)	(∞, ∞)	Yes
4.	Configuring, front–back	FRONT–BACK	Gesture 21	(Deictic, emblem, deictic)	(∞, ∞)	Yes
5.	Configuring, near–far	NEAR–FAR	Gesture 31	(Deictic, emblem, deictic)	(∞, ∞)	Yes
6.	Configuring, contact	CONTACT	Gesture 27	(Deictic, emblem, deictic)	(∞, ∞)	Yes
7.	Configuring, center–periphery	CENTER–PERIPHERY	Gesture 17	(Deictic, emblem, deictic)	(∞, ∞)	Yes
8.	Configuring, before–after	PATH	Gesture 19 $\circ$ ρ	Seq. of deictic gestures	∞	Yes
9.	Linking, directed	PATH	Gesture 19 $\circ$ ρ	Seq. of deictic gestures	∞	Yes
10.	Linking, undirected	LINK	Gesture 15	(Deictic, deictic)	$(1, 1)$	Yes
11.	Rating, absolutely	CONTAINER	Gesture 11 $\circ$ ρ_1 $\circ$ ρ_2	Deictic	1	Yes
12.	Rating, relatively	UP–DOWN	Gesture 23 $\circ$ ρ_1 $\circ$ ρ_2	(Deictic, emblem, deictic)	(∞, ∞)	Yes
13.	Attributing, segments	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	No
14.	Attributing, segment shapes	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	No
15.	Attributing, segment textures	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	No
16.	Attributing, configurations	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	Yes
17.	Attributing, links	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	Yes
18.	Attributing, ratings	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	Yes
19.	Attributing, attributions	CONTAINER	Gesture 11 $\circ$ ρ	Deictic	1	Yes

The numbers by which gestures are identified refer to the corresponding figures. $\circ$ denotes the operation of concatenation. ρ is a variable that denotes a gestural radical. Column 6 accounts for the complexity of argument identification (excluding emblems for separating between in and out arguments). Column 7 and 8 specify whether the predication is intramedial (by focusing on a single image) or intermedial (by covering different images)

themselves—once more by means of radicals. By gesturing the radical, WikiNect can highlight the corresponding domain of objects to narrow down the choice. Attributions are the means of forming complexes of (possibly discontinuous) image segments that do not necessarily belong to the same segment. In this way, the classification of segments by their bottom-up attribution corresponds to the top-down operation of segmentation. Thus, as in the case of ratings, attributions are regarded as a special kind of classification: An object is said to be classified by a certain attribute if it is said to belong to the class of objects that share this attribute. By means of radicals, it is possible to further distinguish whether the attribution regards the shape or the texture of a segment (Lines 14–15 in Table 1). What is left unspecified in this model is the attribute itself: A segment can be said to depict a person, a group of persons can form a team, a link can denote a kinship relation etc. Obviously, the range of such attributions relates to Requirement 2 in that it goes beyond gestural writing in the narrow sense. However, this is the point where it is possible to utilize another feature of image schemata, that is, their role in the formation of metaphors [29, p. 283]. This can be exemplified as follows: In order to structure a conceptual, non-perceptional description of an image, one can name the target domain (say, *power relations*), select the structuring image schema (e.g., UP–DOWN) and perform a mapping of arguments (e.g., onto *(being) up* now in the sense (*being*) *powerful*). In this way, one can state not only that an image segment is located above another one, but also that it is more luminous, or that the depicted person is more powerful or what else is meant by *up* according to the target domain. Image schemata are blueprints of conceptual structuring—whether directly grounded in spatio-temporal experience or not. An approach that extends Model 3 in this way will be called *Model 4* of gestural writing. Note that while Model 3 relies on the "literal meaning" of image schemata, Model 4 additionally allows for their metaphorical use in the sense of Lakoff [29]. However, because of the range of possible target domains, their selection cannot be gestured iconically. Thus, Model 4 falls back onto symbolic text processing of selecting predicates according to a predicate palette or any other linguistic specification of predicates.

Before the expressiveness of Model 3 and 4 can be finally specified, one needs to come back to Examples (2)–(3) to show how they can be handled by Model 4. In case of statement (2-a), one can select the CONTAINER schema before selecting the target domain of metaphorization (i.e., *luminosity*) by means of symbol processing (e.g., using a virtual keyboard). Finally, if being segmented before, the argument-forming area (showing Rembrandt's face) has to be dragged and dropped over the container. Otherwise, a segmentation precedes the drag and drop operation. Likewise, in case of the statement (2-b), the UP–DOWN schema can be used such that *up* is interpreted to mean *more luminous*. Finally, each segment except the one depicting Rembrandt's face is mapped onto the schema's down-slot. Statement (3) is more difficult. It is assumed that the statements (3-a)–(3-c) have been written by analogy to (2-a). In case of statement (3-d), the SOURCE–PATH–GOAL schema is selected before specifying *range* to be the target domain. Then, a path is drawn from the area of darkest gray into the direction of the area of darkest brown to span the range. Finally, in case of statement (3-e), the range is categorized as configured before to be restraint. That is, the CONTAINER schema is selected, followed by the radical for denoting configurations, the target domain (i.e., *restraint*) is specified and the assignment is made by deictically identifying the latter range, dragging and dropping it over the focal container.

In sum, Model 4 exemplifies an interplay of iconic and deictic gestures that in conjunction with symbolic operations of predicate selection facilitate the gestural expression of propositions thereby introducing gestural writing as a novel means of HCI. The question about the degree of expressiveness that it shares with verbal communication is tackled next.

5.2 On the expressiveness of gestural writing by example of image descriptions

Currently, there is no gold standard that allows for rating the expressiveness of gestural writing. However, since, its present object area is image description, the authors' can draw on studies that classify complexity levels of this task. More specifically, it is asked which complexity level of image description is reached by gestural writing.

Several classifications related to this task exist in related literature. A recent overview is given by Benson [4]. The approach of Hollink et al. [20] is utilized since it explored the frequency distribution of different tasks of image description experimentally. In this way, one can get an insight into the effectiveness of gestural writing as defined so far. Starting from an integration of related models (including the pyramidal model of syntactic and semantic levels of image description of Jaimes and Chang [22]), the model of Hollink et al. [20] distinguishes three levels:

1. The description at the *non-visual level* relates to metadata of images, their creators, material, locations etc. Since on this level Hollink et al. [20] include intermedial relations, a first coverage by gestural writing is noticed, which uses the LINK schema or the SOURCE–PATH–GOAL schema to map undirected or directed relations of this sort.
2. The syntactic description of images on the *perceptual level* basically includes the color, shape, texture and

composition of image segments and related visual characteristics that can be described with little recurrence to world knowledge.

3. The semantic description of images on the *conceptual level*: This level is based on descriptions of conceptual objects and their (partly spatio-temporal, event-based) relations. Since it relies on interpretational, meaning-related objects, it requires the full range of an open semantic universe and, thus, is in conflict with Requirement 2 (above).

Evidently, the working area of Model 3 is the syntactic or perceptual level. Apart from intermedial relations, the non-visual level is not addressed by gestural writing as specified so far. In any event, the semantic or conceptual level is out of reach of Model 3. In light of the frequency distribution of operations of image description explored by Hollink et al. [20], this means that gestural writing according to Model 3 is underrepresented in that it focuses on <12 % of these operations.

The semantic level is only reached, though not covered, by Model 4, but at the price of symbolic operations of selecting non-visual, non-spatio-temporal predicates. However, in this way, Model 4 addresses about 87 % of the operations of image descriptions as counted by Hollink et al. [20]. This is only possible by a metaphorical use of image schemata. As these schemata are anchored in gestural writing by means of iconic gestures (see Table 1), the paper paves the way for a gestural adoption of image descriptions on the conceptual level. In sum, a trade-off between the iconicity and indexicality of gestural writing on the one hand and its expressiveness on the other is stated: The more one relies on iconic or indexical means of writing, the less expressive the model and, vice versa, expressiveness on the conceptual level is only reached with symbolic means. In spite of this negative finding, a way to reduce this symbolic load has been found. This has been done with the help of image schemata that allow for mapping predicates in an open semantic universe to a small range of structure providing spatio-temporal predicates.

6 Conclusion

HCI interface design strives after easy handling. The most intuitive forms of interactions are known to be iconic and indexical means. The present paper provides a starting point for fathoming this common HCI view in a semiotic perspective. The rationale of this account relates to answering the following question: *Given the constraining frame of reference of an application scenario (image descriptions, in this case), how much of the symbolic realm of this scenario can be reduced to more direct pre-symbolic interactions?* In order to give an answer in the framework of image descriptions, gestural writing is introduced as a means for such a pre-symbolic communication. Gestural writing is addressed in a bottom-up approach, by relating propositional acts from the given domain of application to more abstract image schemata. The most advanced Models 3 and 4 of gestural writing presented here function nearly exclusively in iconic and indexical terms. In this way, they reach the level of conceptual image descriptions—including the perceptual level that is already reached by Model 3. In any event, Model 4 still needs conventionalized, symbolic elements.

Capturing a broader range of predicates seems to be possible, but only at the expense of ever finer image schemata and, hence, more complex and more artificial gestural representations. The more complex and less natural the gestures, the more difficult to learn for users and the more difficult to track for the Kinect system. This observation leads to a couple of consequences that have to be addressed in future work:

- *What iconic representations are there at all?* A differentiation of the iconic mode beyond mere resemblance has to be given in terms of a theory of signs like that of the philosopher and semiotician Charles Sanders Peirce. Such a differentiation would deliver a better picture of the gestures hypothesized to fit iconic requirements very well.
- The fanning out of a variety of gestures as needed to provide representations for ever more fine-grained schemata are naturally limited by the human anatomy. Certain kinds of movements are simply not possible. As an example, recall the *center–periphery* gesture from Fig. 17. It is impossible to draw a full circle there, since this trajectory is blocked by the eventual crossing of arms. On the other hand, more subtle movements may not be recognized by the tracking system. That is, an analysis of gestural writing needs to be complemented by a provision of clear-cut and distinct gestures that still keep an iconic or indexical kernel.
- A single image schema can be instantiated by more than one gesture, even by more than one iconic gesture. Take, for example, the *container* gesture from Fig. 11. The container can be depicted by various handshapes, for instance, by a bent hand or by two hands forming a closure gestalt. It is hard to tell such gestures apart in iconic terms exclusively—they appear to be largely equivalent in this respect. However, the *production* of these gestures is not equally comfortable. Therefore, it seems reasonable to assume that gestural writing has to include an assessment of the "morphological simplicity" of gestures at least as a selection mechanism for gestures that are on a par otherwise.

Exploring the options of gestural writing as a pre-symbolic strategy for natural HCI interactions needs to be complemented by a couple of considerations like the ones identified above. Accordingly, in partly already started future work, gesture vocabularies designed for a certain set of tasks are evaluated in terms of their kinematic and semantic convenience in an experimental machine learning setting. However, the conceptual gauging of task-related propositional acts in terms of more abstract cognitive structures for understanding provides a starting point for a principled assessment of the complexity and intuitiveness of gesture-based interfaces in HCI. In this sense, by Model 4, a first instantiation of gestural writing is elaborated by means of deictic and iconic gestures.

Acknowledgments Financial support by the program *Wandel gestalten!* of the *Heinz Nixdorf Stiftung* and the *Stifterverband* as well as by *Microsoft Germany* is gratefully acknowledged. The authors also thank the anonymous reviewers for their helpful comments.

References

1. Asir, A., Creech, B., Homburg, T., Hoxha, N., Röhrl, B., Stender, N., Uslu, T., Wiegand, T., Kastrati, L., Valipour, S., Akemlek, D., Auth, C., Hemati, A., Korchi Said Omari, S., Schöneberger, C.: Practical course WikiNect at the Department of Computer Science, Goethe University Frankfurt (2013). Praktikumsseminar
2. Bartoli, G., Del Bimbo, A., Faconti, M., Ferracani, A., Marini, V., Pezzatini, D., Seidenari, L., Zilleruelo, F.: Emergency medicine training with gesture driven interactive 3d simulations. In: Proceedings of the 2012 ACM Workshop on User Experience in e-learning and Augmented Technologies in Education. UXeLATE '12, pp. 25–30. ACM, New York, NY (2012)
3. Barwise, J., Perry, J.: Situations and Attitudes. MIT Press, Cambridge (1983)
4. Benson, A.C.: Relationship analysis of image descriptions: an ontological, content analytic approach. Ph.D. thesis, University of Pittsburgh (2011)
5. Bernstein, M.: Can we talk about spatial hypertext? In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, HT '11, pp. 103–112. ACM, New York, NY (2011) doi:10.1145/1995966.1995983
6. Bouissac, P.: The study of metaphor and gesture: a critique from the perspective of semiotics. In: Cienki, A.J., Müller, C. (eds.) Metaphor and Gesture, pp. 277–282. John Benjamins, Amsterdam (2008)
7. Brennan, S.E.: The grounding problem in conversations with and through computers. In: Fussell, S.R., Kreuz, R.J. (eds.) Social and Cognitive Psychological Approaches to Interpersonal Communication, pp. 201–225. Lawrence Erlbaum, Hillsdale, NJ (1998)
8. Bundesverband Museumspädagogik e.V. (ed.): Qualitätskriterien für Museen: Bildungs- und Vermittlungsarbeit. Holzer Druck und Medien, Weiler i. Allgäu (2008)
9. Chapinal Cervantes, J., Vela, F.L.G., Rodríguez, P.P.: Natural interaction techniques using Kinect. In: Proceedings of the 13th International Conference on Interacción Persona-Ordenador, INTERACCION '12, pp. 14:1–14:2. ACM (2012)
10. Clausner, T.C., Croft, W.: Domains and image schemas. Cogn. Linguist. **10**(1), 1–31 (1999)
11. Cochran, Z.R.: The bit dome: creating an immersive digital environment with a Kinect-based user interface. J. Comput. Sci. Coll. **29**(2), 191–198 (2013)
12. Community on Education, American Association of Museums: excellence in practice—museum education principles and standards. www.edcom.org/Files/Admin/EdComBookletFinalApril805.pdf (2005)
13. de Ruiter, J.P.: Postcards from the mind: the relationship between speech, imagistic gesture, and thought. Gesture **7**(1), 21–38 (2007)
14. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica **1**(1), 49–98 (1969)
15. Fillmore, C.J.: The case for case. In: Bach, E., Harms, R.T. (eds.) Universals in Linguistic Theory, pp. 1–88. Holt, Rinehart and Winston, New York (1968)
16. Foss, J.G., Cristea, A.I.: The next generation authoring adaptive hypermedia: using and evaluating the mot3.0 and peal tools. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, HT '10, pp. 83–92. ACM, New York, NY (2010). doi:10.1145/1810617.1810633
17. Gasperetti, B., Milford, M., Blanchard, D., Yang, S.P., Lieberman, L., Foley, J.T.: Dance dance revolution and eyetoy kinetic modifications for youths with visual impairments. J. Phys. Educ. Recreat. Dance **81**(4), 15–55 (2010)
18. Goschler, J.: Embodiment and body metaphors. metaphorik.de 09, 33–52 (2005). http://www.metaphorik.de/09/
19. Haser, V.: Metaphor, Metonymy, and Experientialist Philosophy: Challenging Cognitive Semantics. Mouton de Gruyter, Berlin (2005)
20. Hollink, L., Schreiber, A.T., Wielinga, B.J., Worring, M.: Classification of user image descriptions. Int. J. Hum. Comput. Stud. **61**(5), 601–626 (2004)
21. Inceoglu, M.R.: WikiNect: Bewegungsunterstützte Texttechnologie. B.A. thesis, Goethe University (2013)
22. Jaimes, A., Chang, S.F.: A conceptual framework for indexing visual information at multiple levels. In: IS&T/SPIE Internet Imaging, vol. 3964 (2000)
23. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: a survey. Comput. Vis. Image Underst. **108**(1–2), 116–134 (2007). doi:10.1016/j.cviu.2006.10.019
24. Johnson, M.: The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. University of Chicago, Chicago (1987)
25. Karam, M., Schraefel, M.C.: A taxonomy of gestures in human computer interactions. Technical report, University of Southampton (2005). http://eprints.soton.ac.uk/261149/
26. Kendon, A.: Gesticulation and speech: two aspects of the process of utterance. In: Key, M.R. (ed.) The Relationship of Verbal and Nonverbal Communication, Contributions to the Sociology of Language, vol. 25, pp. 207–227. Mouton, The Hague (1980)
27. Kendon, A.: Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge, MA (2004)
28. Kertész, A., Rákosi, C.: Cyclic vs. circular argumentation in the conceptual metaphor theory. Cogn. Linguist. **20**(4), 703–732 (2009). doi:10.1515/COGL.2009.030
29. Lakoff, G.: Women, Fire, and Dangerous Things: What Categories Reveal About the Mind. University of Chicago Press, Chicago (1987)

30. Lausberg, H., Sloetjes, H.: Coding gestural behavior with the NEUROGES–ELAN system. Behav. Res. Methods **41**(3), 841–849 (2009)
31. Leuf, B., Cunningham, W.: The Wiki Way: Quick Collaboration on the Web. Addison Wesley, Boston (2001)
32. Lücking, A.: Zugl. Ikonische Gesten. Grundzüge einer linguistischen Theorie. De Gruyter, Berlin (2013). Zugl. Diss. Univ. Bielefeld
33. Lücking, A., Mehler, A., Menke, P.: Taking fingerprints of speech-and-gesture ensembles: approaching empirical evidence of intrapersonal alignment in multimodal communication. In: LonDial 2008: The 12th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL), pp. 157–164. King's College London (2008)
34. Marquardt, Z.: Beira, J.A., Em, N., Paiva, I., Kox, S.: Super mirror: a kinect interface for ballet dancers. In: CHI '12 Extended Abstracts on Human Factors in Computing Systems. CHI EA '12, pp. 1619–1624. ACM, New York, NY (2012)
35. Marsh, T.: An iconic gesture is worth more than a thousands words. In: Proceedings of the International Conference on Information Visualisation. IV '98, pp. 222–223. Washington, DC (1998)
36. McGlone, M.S.: Concepts as metaphors. In: Glucksberg, S. (ed.) Understanding Figurative Language: From Metaphors to Idioms, pp. 90–107. Oxford University Press, Oxford (2001)
37. McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. Chicago University Press, Chicago (1992)
38. Mehler, A., Lücking, A.: WikiNect: towards a gestural writing system for kinetic museum wikis. In: Proceedings of the International Workshop On User Experience in e-Learning and Augmented Technologies in Education, UXeLATE 2012, pp. 7–12 (2012). Workshop held in Conjunction with ACM Multimedia
39. Mittelberg, I.: Peircean semiotics meets conceptual metaphor: iconic modes in gestural representations of grammar. In: Cienki, A.J., Müller, C. (eds.) Metaphor and Gesture, pp. 115–154. John Benjamins, Amsterdam (2008)
40. Morelli, T., Folmer, E.: Real-time sensory substitution to enable players who are blind to play video games using whole body gestures. In: Proceedings of the 6th International Conference on Foundations of Digital Games. FDG '11, pp. 147–153. ACM, New York, NY (2011)
41. Müller, C.: Redebegleitende Gesten. Kulturgeschichte - Theorie - Sprachvergleich, Körper - Kultur - Kommunikation, vol. 1. Berlin Verlag, Berlin (1998)
42. Panger, G.: Kinect in the kitchen: testing depth camera interactions in practical home environments. In: CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12, pp. 1985–1990. ACM (2012)
43. Peirce, C.S.: Collected Papers of Charles Sanders Peirce, vol. II. Harvard University Press, Cambridge, MA (1965). Repr. from 1932
44. Poggi, I.: From a typology of gestures to a procedure for gesture production. In: Wachsmuth, I., Sowa, T. (eds.) Gesture and Sign Language in Human–Computer Interaction, Lecture Notes in Computer Science, vol. 2298, pp. 158–168. Springer, Berlin (2002). doi:10.1007/3-540-47873-6-16
45. Quek, F.K.H., McNeill, D., Bryll, R.K., Duncan, S., Ma, X.F., Kirbas, C., McCullough, K.E., Ansari, R.: Multimodal human discourse: gesture and speech. ACM Trans. Comput. Hum. Interact. **9**(3), 171–193 (2002)
46. Radkowski, R., Stritzke, C.: Interactive hand gesture-based assembly for augmented reality applications. In: ACHI 2012, The Fifth International Conference on Advances in Computer–Human Interactions pp. 303–308 (2012)
47. Rector, K., Bennett, C.L., Kientz, J.A.: Eyes-free yoga: an exergame using depth cameras for blind and low vision exercise. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '13, pp. 12:1–12:8. ACM, New York, NY (2013)
48. Rohrer, T.: Image schemata in the brain. In: Hampe, B. (ed.) From Perception to Meaning. Image Schemas in Cognitive Linguistics, Cognitive Linguistics Research, vol. 29, pp. 165–196. Mouton de Gruyter, Berlin (2005)
49. Schöfegger, K., Körner, C., Singer, P., Granitzer, M.: Learning user characteristics from social tagging behavior. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT '12, pp. 207–212. ACM, New York, NY (2012). doi:10.1145/2309996.2310031
50. Searle, J.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University, London (1969)
51. Seroussi, Y., Bohnert, F., Zukerman, I.: Personalised rating prediction for new users using latent factor models. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, HT '11, pp. 47–56. ACM, New York, NY (2011). doi:10.1145/1995966.1995976
52. Solis, C., Ali, N.: An experience using a spatial hypertext wiki. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, HT '11, pp. 133–142. ACM, New York, NY (2011). doi:10.1145/1995966.1995986
53. Streeck, J.: Depicting by gesture. Gesture **8**(3), 285–301 (2008). doi:10.1075/gest.8.3.02str
54. von Ahn, L.: Games with a purpose. Computer **39**(6), 92–94 (2006). doi:10.1109/MC.2006.196
55. Wechsung, I., Engelbrecht, K.P., Kühnel, C., Möller, S., Weiss, B.: Measuring the quality of service and quality of experience of multimodal human–machine interaction. J. Multimodal User Interfaces **6**, 73–85 (2012). doi:10.1007/s12193-011-0088-y