Comparing Hand Gesture Vocabularies for HCI

Alexander Mehler, Tim vor der Brück, and Andy Lücking

Text Technology Lab, Department of Computer Science and Mathematics,
Goethe-University Frankfurt am Main

Abstract HCI systems are often equipped with gestural interfaces drawing on a predefined set of admitted gestures. We provide an assessment of the fitness of such gesture vocabularies in terms of their learnability and naturalness. This is done by example of rivaling gesture vocabularies of the museum information system WikiNect. In this way, we do not only provide a procedure for evaluating gesture vocabularies, but additionally contribute to design criteria to be followed by the gestures.

1 Motivation

Hand gestures are of great interest for HCI applications, since they are considered to help "to develop more natural and efficient human-computer interfaces." [1] There are two kinds of prevalent HCI gestures: *manipulators* and *semaphores* [2]. Manipulators are actions that manipulate some entity provided by the display – for instance, pushing a button or moving a slider. Therefore, manipulators are largely driven by the displayed entity and its functionality. This "tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated" [2, p. 172] is not a defining feature of semaphores. Rather, semaphoric gestures are hand/arm forms that are organized as a predefined, often stylized vocabulary, or lexicon [2, p. 173]. Such gesture vocabularies can be designed in a better or worse way. Semaphores are considered to be better, if they are more "intuitive" or "motivated". Motivatedness is accomplished if the form (hand shape, movement trajectory) of a gesture "imitates the referent by selecting one or more of its visually perceivable features" [3, 49]. In other words: intuitive gestures resemble their object, they are *iconic*. However, it is well known now that iconic gestures do not signify or refer on their own. Rather, other means are required for establishing signification, for instance, a conventional one [4]. Conventionality involves arbitrariness that has to be mastered by learning. Of course, users favor gesture vocabularies that can be learned easily [5, p. 33]. Accordingly, a second dimension for evaluating sets of semaphores has to be their learnability.

Both lines of assessing the fitness of gesture vocabularies have been pursued in previous research by different methodologies, for example:

- the naturalness of gesture vocabularies has been investigated by [6] by means of user studies;
- the learnability of semaphores (including an empirically specified intuitiveness index) have been studied as an analytical optimization problem by [7].

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 81–92, 2014.

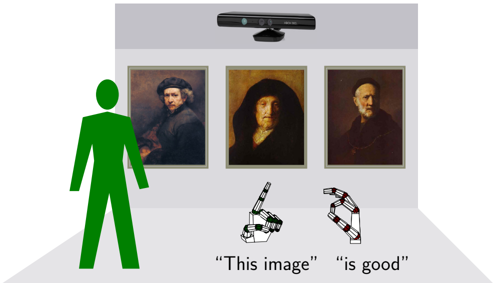

Fig. 1. WikiNect application scenario: rating of an image (taken from [9])

one.handed.gesture		
Two.Handed.Conf	0	
Mov.Relative	0	
RH	*right.hand*	
	Handshape.B	
	Handshape.Mov	0
	Handshape.Path	0
	Palm.Orient	PAB
	Palm.Mov	0
	Palm.Path	0
	BoH.Orient	BUP
	BoH.Mov	0
	BoH.Path	0
	Wrist.Loc	CC
	Wrist.Dist	D-O
	Wrist.Mov	MDR>MUR
	Wrist.Path	LINE>LINE
	Mov.Extent	M>L
	Temporal.Sequence	0

Fig. 2. Representation of Checkmark gesture from Table 3

The latter work deals with static hand configurations from one robotic arm control vocabulary. The present paper further develops optimization procedures for gesture vocabularies, mainly in two respects:

1. firstly, in addition to static gesture, also dynamic gestures are accounted for;
2. secondly, evaluation is not only based on one gesture vocabulary, but is carried out as a comparison between different sets of semaphores.

The testing environment for the comparison of gesture vocabularies is the WikiNect system [8] (see also `www.hucompute.org/ressourcen/wikinect`). WikiNect is a platform for the gestural writing of wikis in the context of museums. Using the Kinect technology, WikiNect allows for a non-contact, gesture-based segmentation, linkage, attribution and rating of (segments of) images. As an on-site museum information system, WikiNect aims at enabling museum visitors to describe, evaluate and comment images of the corresponding exhibition. In Figure 1 (taken from [9]) a typical WikiNect application scenario is given where a user selects an image by means of a pointing gesture and appreciates it using a semaphoric, codified "OK" gesture.

Being an HCI application that is addressed to the diverse audience of museum visitors, WikiNect itself has an interest in natural and learnable gesture-based interactions. Accordingly, the gesture vocabularies to be evaluated are taken from two prototype implementations of WikiNect [10,11]. To this end, Section 2 describes the gesture vocabularies in conjunction with a subset of tasks accomplished by WikiNect. Section 3 accounts for task-gesture mappings in terms of a quadratic optimization problem. It starts from a quantitative analysis of Wikipedia-based image descriptions which results in a corresponding set of soft constraints. The evaluation rationale and experimentation for assessing gesture

Table 1. Selected tasks accomplished by WikiNect

Navigation Tasks	**Segmentation Tasks**	
Scrolling backward	Select image	Circular segment
Scrolling forward	Segment image	Rectangular segment
Close, back to Main	Save image	Polygonal segment
Undo	Display segments	Free-hand segmentation

Table 2. Spatial expressions partitioned according to three spatial modalities *Direction*, *Relations*, and *Form*

Direction	**Relation**		**Form**
left	above	behind	circle
right	below	through	rectangle
up	by	at	triangle
down	in	on	cornered
front	around	between	bent
back	in front of	along	random
			straight

vocabularies is finally presented in Section 4, while Section 5 provides a concluding discussion.

2 WikiNect Gestures, Tasks and Annotations

The usage of WikiNect is subdivided into a navigation and a segmentation component [8]. Navigation gestures are used for selecting WikiNect's functional modules, while segmentation gestures are operative in the segmentation mode. Table 1 lists 12 of these tasks which have been implemented in two prototype systems according to different design strategies [10,11]. The first prototype, hereafter called WN-1, provides a set of controlling gestures taken from the InkCanvas class of the `.NET` Framework and mapped onto the system's operations [10]. The second prototype, WN-2, can be operated mainly by manipulation gestures (e.g., by pushing buttons that trigger a certain operation) [11].

Any gesture used to implement WikiNect has been represented in terms of spatial predicates. The rationale behind this is to allow for task-gesture mappings: gestures are preferably mapped to tasks with which they share many predicates. In order to obtain a set of spatial predicates, we use the list of the spatial predicates collected by [12, p. 97]. This list has been extended by (1) the directions spanned along the body axes and (2) basic form-related predicates. The spatial predicates are partitioned according to the spatial modalities *direction*, *relation* and *form* – see Table 2. They are used to label both the tasks and the gestures for spatial properties, either quite literally or associatively. Some notes on the application of the predicates:

Table 3. Navigational gestures used in [10] for implementing WikiNect

Gesture Annotation	Image	Movement	Task	Task Annotation	Naturalness
right, straight, along		towards right	Scrolling backward	back, below, left	0
left, straight, along		towards left	Scrolling forward	front, up, right	0
through, cornered, down, up, right		Checkmark, towards down-left, towards up-right	(1) Select image, (2) Segment image, (3) Save image	(1) around, through; (2) in, through; (3) in, random	(1) 0.077; (2) 0.1; (3) 0
right, through, cornered, up, around, above		towards right, upward	Close active window, back to main	back, below, left	0
through, right, cornered, down, around, below		towards right, downward	(1) Undo, (2) Display segments of an image	(1) back, down, left, random; (2) in, around, by, random	(1) 0.033; (2) 0.031

- If a movement comprises a change of direction, it is understood as to run *through* the turning point and the predicate "through" is chosen.
- If a task contains a temporal aspect like *backwards* (i.e., going back in the system's history), three conceptualizations are acknowledged:
  1. Stack – orientation along longitudinal axis ("up", "down");
  2. Tape – orientation along transversal axis ("right", "left");
  3. Gaze – orientation along sagittal axis ("front", "back").
- Closed forms give rise to containment indicated by "in".

We emphasize that the annotation so far has the status of a working hypothesis. We aim at demonstrating that our approach is feasible and provides useful results without claiming that the predicate list is the only possible one.

For illustration, the description and annotation of gestures and tasks of WN-2 is given in Tables 3 and 4. The columns "Movement" and "Image" contain a shorthand and a pictorial representation of the gestures. The column "Naturalness" shows the naturalness index calculated according to the procedure explained in Section 4.1. To make the gestures' forms objects of quantitative analyses, they are coded according to the kinematic-oriented representation format of [13] – see Figure 2 for an example. Based on text-based representations of this kind, we apply distance measures in optimizing task-gesture mappings.

Table 4. Segmentation gestures used in [10] for implementing WikiNect

Gesture Annotation	Image	Movement	Task	Task Annotation	Naturalness
circle, around, bent		Circle	Cut out circular segment	circle, bent, in, around	0.18
around, rectangle, cornered		Rectangle	Cut out rectangular segment	in, cornered, around, rectangle	0.15
around, triangle, cornered		Triangle	Cut out polygonal segment	triangle, in, around, cornered	0.15
through, left, right, down, cornered, between		Towards down-left, towards down-right	Activate free-hand segmentation	random, in, around, circle, rectangle, triangle	0

3 Towards Optimal Task-Gesture Mappings

The task of image description is schematized to a certain degree [14]. WikiNect deals with four such routinized tasks: *rating*, *segmenting*, *linking* and *attributing* images (e.g., with information about painters or techniques). Our aim is to find gestural representations of theses tasks so that users can make image descriptions by using WikiNect, that is, by *gestural writing* [9]. A naïve way to realize this would be to select from an artificial lexicon of prespecified gestures. The problem is rather how to justify any mapping of image description tasks onto gestures. An iconic gesture, for example, is a natural candidate to manifest a gestalt-related image description, while a deictic gesture is a better candidate for selecting images on the screen.

Our approach to solve this problem is twofold: firstly, we analyze Wikipedia as the biggest sample of image descriptions to learn about the frequency distributions of the actions involved in such descriptions. Secondly, we utilize this information to derive constraints that any procedure of gesture selection should fulfill to provide both efficiently producible and learnable gestures for gestural writing. This approach follows a twofold optimization criterion: we select gestures for actions of image descriptions such that the more frequent the action the more easily producible the gesture while preserving a certain amount of discriminability (i.e., learnability) among gestural manifestations of different actions.

Information about the frequency distributions of image description tasks is not directly accessible for lack of large-scale annotations of corresponding speech acts. However, the English Wikipedia offers a range of data to approach this information. To learn about the frequency distribution of linking images, for example, we can explore hyperlinks between articles about these images (see Table 5 for a

Table 5. Statistics of image description articles in the English Wikipedia

Attribute	Value
articles	2,862
instances of painting/artwork template	2,926
links among the 2,862 articles	62,725
corpus size	14.7 MB
average size (per article)	5.3 KB
date of extraction	2014=2012 November 1, 2014

statistics of the underlying corpus; see Figure 3 for the resulting distribution (distributions have been shifted by one, to account for zero frequencies)). Likewise, to get information about the frequency distribution of image attributions, we explore every instance of `Template:Infobox_artwork`[1] (Figure 4). Next, since there is no matching template for segmenting images, we need to assess the corresponding frequency distribution indirectly. This is done by exploring the frequency distribution of section headers like `Composition`, `Analysis` or `Details` within the corpus of image articles (Figure 5). Likewise, because of the lack of directly accessible ratings of images, we explore the ratings of their corresponding articles (as manifested by the `Rate this page`-section). In this way, we approximate a frequency distribution of image-related ratings (Figure 6). As can be seen by Figures (3–6), each of the four tasks (*linking*, *attributing*, *segmenting* and *rating*) results in a power-law-like frequency distribution being reminiscent of Zipf's law of least effort [15]. Only a couple of images is, for example, linked to many other images while most images are linked only once (Figure 3). Likewise, there is a small set of predominant attributes while most attributes are rarely used if at all. Further, the frequency distribution of section headers shows a small set of predominant sections (Figure 5) that leave behind a huge set of rarely used ones: apart from conventional sections in Wikipedia (e.g. `References` or `External links`), the former set is exemplified by headers like `Artist`, `Description` and `Composition`. That is, when writing about the content of images, Wikipedians follow a power law according to which they prefer a small range of topics of highest probability. Analogously, the distribution of the numbers of ratings strictly follows a power law (Figure 6) in any of the four dimensions considered by Wikipedia: a few images have many ratings while most images have few ratings or none at all.

In sum, image descriptions follow a highly skewed distribution such that the frequencies of the underlying actions decay according to a power law. Thus, when looking for gestural manifestations of such actions we can follow the example of natural languages [15]: the more frequent an action the simpler its manifestation should be. Since we need to manifest different actions simultaneously, we additionally need to preserve discriminability among neighboring ranks in the

[1] We also explore `Template:Infobox_Painting` which redirects to `Template:Infobox_artwork`.

frequency distribution of gestural manifestations. As a rule of thumb: *optimizing along the criterion of least effort should not happen at the expense of discriminability and thus learnability among highly frequent gestures.* In what follows, we represent this finding in terms of a *quadratic integer programming problem* whose solution leads to the optimal task-gesture mapping – subject to the operative constraints (number of tasks, gesture repertoire etc.).

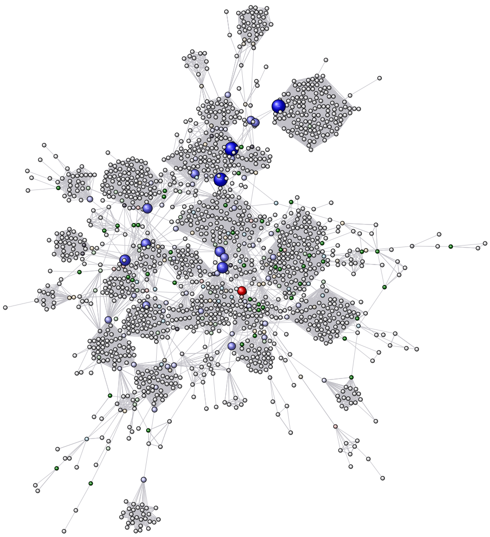

Fig. 3. The largest weakly connected component of the article graph of image descriptions in the English Wikipedia that covers 56% of the descriptions. The distribution of the node degrees of this graph follows a power law with exponent 1.55 (according to [16], we fit the complementary cumulative distribution $P(X \geq x)$ that yields an exponent of 0.55 ($\overline{R}^2 = 94\%$); according to [17] this corresponds to an exponent of 1.55 in terms of $P(X = x)$. The same procedure is applied in all fittings).

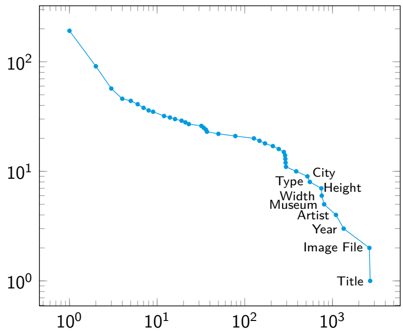

Fig. 4. 1-shifted complementary cumulative distribution of attribution templates for images (exp. 1.26 (2.26), $\overline{R}^2 = 97.7\%$)

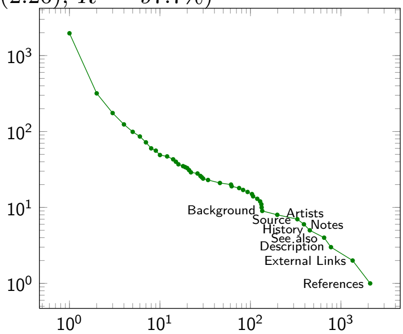

Fig. 5. Complementary cumulative distribution of section headers (exponent 1.55 (2.55), $\overline{R}^2 = 99\%$)

Generally speaking, a quadratic integer programming problem requires all decision variables to be integer, while its constraints are required to be linear and the objective function to contain a quadratic term. To reformulate this in terms of gesture modeling, we proceed as follows: let n be the number of gestures and m the number of tasks. Assume that gestures and tasks are all numbered so that the set of tasks is given by $T = \{t_1, \ldots, t_n\}$ and the set of gestures by $G = \{g_1, \ldots, g_m\}$. The decision variables in this mapping problem

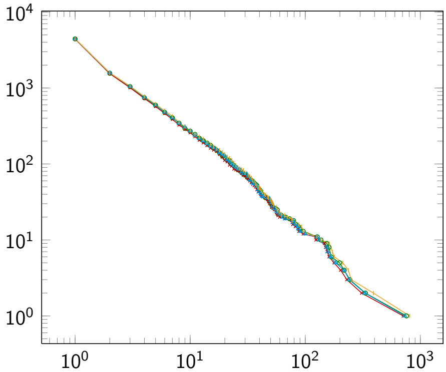

Fig. 6. 1-shifted complementary cumulative distributions of ratings show four distributions of the rating template (*trustworthy* (green circle, exp. 0.7 (1.7), $\overline{R}^2 =$ 99%), *well-written* (orange bars, exp. 0.72 (1.72), $\overline{R}^2 = 99\%$), *objective* (red crosses, exp. 0.7 (1.7), $\overline{R}^2 = 99\%$), and *complete* (blue triangles, exp. 0.7 (1.7), $\overline{R}^2 = 99\%$))

Table 6. Frequency distribution of tasks by predicates

Task	Freq.	Perc. %
Circular segment	1,180	10.37
Close, back to Main	719	6.32
Display segments	1,399	12.29
Free-hand segmentation	1,189	10.45
Rectangular segment	1,172	10.3
Save image	1,121	9.85
Scrolling backward	719	6.32
Scrolling forward	827	7.27
Segment image	1,132	9.95
Select image	62	0.54
Polygonal segment	1,171	10.29
Undo	691	6.07
Sum	11,382	100

are binary features x_{ij} that are 1 if gesture g_i should be mapped to task t_j and zero otherwise. A hard constraint is to require that each task is always mapped to a single gesture, i.e., synonymous gestures and not-assigned tasks are not allowed. We formalize this by means of equality constraints:

$$\sum_{i=1}^{n} x_{ij} = 1 \text{ for } j = \{1, \ldots, m\} \; ; \tag{1}$$

Since the number of gestures exceeds the number of taks, some gestures have to be polysemous and are assigned to several tasks. For the gestures, we only require that each gesture is assigned to at least one action:

$$\sum_{j=1}^{m} x_{ij} \geq 1 \text{ for } i = \{1, \ldots, n\} \tag{2}$$

In addition to hard constraints, three soft constraints are encoded into the objective function:

1. *The simpler the gesture, the more frequent the action to which it is mapped.*
2. *The more frequent an action, the more motivated the gesture mapped onto it.* Since the mapping of gestures to actions has to be memorizable, it should be motivated as much as possible (as explained in Section 4.1).
3. *The more frequent two actions, the easier the discriminability of their gestural manifestations.*

We represent Constraint 1 and 2 by a linear model and Constraint 3 by a squared term as part of the objective function. Given two sets of tasks and gestures (Section 2), an assessment of the motivation of any candidate task-gesture-relation (Section 4.1), a frequency distribution of tasks (Section 4.2), and a measure of the discriminability of gestures (based on their matrix representations – see Figure 2 and [13]), we finally get an optimization problem whose solution, henceforth called **gesture optimizer**, leads to an optimal task-gesture mapping subject to the operative constraints.

4 Experimentation

In this section, we compare two instantiations of the gesture optimizer and contrast them with their corresponding null-models of random task-gesture assignments. To this end, we utilize both implementations of WikiNect (see Sec. 2).

For instantiating the optimizer, we first need to specify two boundary conditions: the motivation of task-gesture relations and the frequency distribution of image description tasks.

4.1 On the Naturalness of Task-Gesture Relations

In order to find an optimal mapping of gestures onto tasks, one needs to know the degree of motivation by which a candidate gesture fits as a manifestation of the tasks. If a user wants to move, for example, something to the left of the display, it is a bad choice to signal this by moving the hand to the right. We provide a simple quantification of this sort of naturalness in terms of bipartite graphs whose bottom mode comprises the candidate gestures and whose top mode is spanned by the tasks under consideration. For any pair $\{g, t\}$ of gestures g and tasks t, an edge occurs in the graph whose initial weight equals the overlap of the predicate descriptions $P(g)$ and $P(t)$:

$$w_1(\{g, t\}) = \frac{|P(g) \cap P(t)|}{\min(|P(g)|, |P(t)|)} \tag{3}$$

Next, we account for diversification in the bipartition. The reason is to prefer unifying task-gesture mappings (in terms of $1:1$ mappings). To see this, think of a system of n tasks, $n \gg 2$, mapped onto one or two gestures. Because of the polysemy of the gestures (as a function of the predicates assigned to them), this system tends to be unnatural: it leads to a semantic overload of the gestures in question. Thus, we re-weight edges as follows (d_v is the degree of vertex v in the bipartition):

$$w_2(\{g, t\}) = w_1(\{g, t\}) \cdot \frac{2}{d_g + d_t} \tag{4}$$

Obviously, a $1:1$-mapping does not alter w_1. Conversely, if the gesture is polysemous or the task is manifested by different gestures, then $w_2 < w_1$. Finally, for any gesture (task), we get a rank order of tasks (gestures) according to their decreasing degree of naturalness. Note that the edge weights are ordinally scaled.

Table 7. Assignments determined by the optimizer for scenarios 1 and 2

Task	Gesture (Scenario 1)	Gesture (Scenario 2)
Scrolling backwards	Left	Grab and drag left
Close window	Left	Grab and drag left
Save Image	Left	Grab and drag right
Display image segments	Right below	Grab and drag right
Scrolling forward	Right	Push forward
Selection	Right above	Push forward
Segment Image	Checkmark	Push forward
Circular segment	Circle	Push forward
Free-hand segmentation	Circle	Push forward
Rectangular segment	Rectangle	Set image point
Polygonal segment	Triangle	Set image point
Undo	Open triangle	Grab and drag left

4.2 Towards a Frequency Distribution of Image Description Tasks

In order to provide a frequency distribution of image description tasks for implementing the *gesture optimizer*, we cannot rely on the Wikipedia data explored in Section 3. The reason is that we focus on the specific task list of WikiNect (see Table 3 and 4). Thus, we alternatively analyze a specialized corpus of image descriptions [18]. The aim is to estimate the probability by which the tasks of Table 1 are conducted in sessions of image description. Since the Wally corpus [18] does not annotate this information and since some of the focal tasks are even not observable in the corpus, we account for this probability indirectly. Following the former sections, we relate tasks and gestures by the predicates they share in their descriptions (see Table 3 and 4). As we map a range of expressions onto these predicates (e.g., *round* and *around* are explored as manifestations of the same-named predicate `around`), the mapping is done by observing the corpus frequencies of the predicates' verbal manifestations. The result of this mapping is shown in Table 6. In contrast to our findings of Section 3, this distribution does not fit a power-law. This may hint at insufficient or even erroneous descriptions of tasks and gestures. For example, though we additionally accounted for multi-word expressions (e.g., *in the front of*), we did not resolve paraphrases of spatial descriptions. Thus, Table 6 has to be understood as a first attempt to estimating the frequencies in questions.

4.3 Results

We tested our approach on two scenarios: given the set of tasks listed in Table 1, the scenarios are distinguished by using the WN-1 and WN-2 set of gestures, respectively. For both scenarios, we determined the optimal assignment for the decision variables by means of the Gurobi optimizer[2] and therefore the optimal mapping

[2] `http://www.gurobi.com`

Table 8. Values of the objective function as determined by the optimizer and the base line method

	Scenario 1	Scenario 2
Optimized value	−2.09	−1.14
Baseline value	−0.98	−0.55

from tasks to gestures that minimizes the objective function (see Table 7 for the optimal mappings).[3]

As a base line, we estimated the expectation value of the objective function by generating 1,000 random assignments of tasks to gestures that fulfill the hard constraints of the optimization problem. The evaluation shows that the optimizer determines assignments for both scenarios for which the objective function values are lower than the base line values (see Table 8 – recall that the lower the objective value the easier to learn and more natural the assignment). Furthermore, the optimal value of the objective function of scenario 1 is below the optimal value of scenario 2, which indicates that scenario 1 is the superior one in terms of learnability and naturalness. Since the number of gestures exceeds the number of tasks in both scenarios, some gestures have to be assigned to more than one task. As can be seen in Table 7, for instance, the *Circle* gesture from scenario 1 is assigned to both the tasks *circle* and *free-hand segmentation*, since both tasks can be chosen in the same context. Thus, the gesture *Circle*, which intuitively is strongly related to circular segmentation mode, gets ambiguous under this assignment. This observation hints at context as a further parameter for improving our model in future work.

5 Conclusion

Based on the notions of *learnability* and *naturalness*, we provide the *gesture optimizer*, a method to assess the fitness of HCI gesture vocabularies to a set of tasks. Optimization is expressed as a quadratic integer programming problem sensitive to a number of constraints. The method is tested in a gesture vocabulary comparison of two WikiNect implementations. Given frequency information of the tasks, a discriminability order between the gestures and a naturalness index based on spatial annotations for gesture-task mappings, we found that the gesture optimizer not only distinguishes gesture vocabularies from a random baseline, but also ranks the vocabularies in the intuitively correct way. Thus, in order to provide an assessment for HCI gestures, the gesture optimizer fuses information and considerations from different sources. Not all of these sources are fully developed yet. However, even given these conditions, we could show that *naturalness*, *frequency* and *learnability* are effective design criteria for devising good HCI gesture vocabularies. This result shows that existing vocabularies

[3] For the second scenario, the optimizer was able to determine the optimal value, for the first scenario we used the best solution found before reaching a time limit.

(think, e.g., of touch gestures!) can be evaluated and, possibly, improved. The gesture optimizer also delineates criteria for designing new vocabularies, so that the method proposed here has many practical applications and provides a test bed for further studies on the fitness of HCI gestures.

References

1. Pavlovic, V.I., Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human-computer interaction: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 677–695 (1997)
2. 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 Transactions on Computer-Human Interaction 9(3), 171–193 (2002)
3. Poggi, I.: Iconicity in different types of gestures. Gesture 8(1), 45–61 (2008)
4. Burks, A.W.: Icon, index, and symbol. PPR 9(4), 673–689 (1949)
5. Baudel, T., Beaudouin-Lafon, M.: Charade: Remote control of objects using free-hand gestures. Communications of the ACM 36(7), 28–35 (1993)
6. Grandhi, S.A., Joue, G., Mittelberg, I.: Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In: Proceedings of SIGCHI, pp. 821–824 (2011)
7. Stern, H., Wachs, J., Edan, Y.: A method for selection of optimal hand gesture vocabularies. In: Sales Dias, M., Gibet, S., Wanderley, M.M., Bastos, R. (eds.) GW 2007. LNCS (LNAI), vol. 5085, pp. 57–68. Springer, Heidelberg (2009)
8. Mehler, A., Lücking, A.: WikiNect: Towards a gestural writing system for kinetic museum wikis. In: Proceedings of UXeLATE, pp. 7–12 (2012)
9. Mehler, A., Lücking, A., Abrami, G.: WikiNect: Image schemata as a basis of gestural writing for kinetic museum wikis. Universal Access in the Information Society (2014) (accepted)
10. Inceoglu, M.R.: WikiNect. BA thesis, Goethe University Frankfurt (2013)
11. 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.: Praktikum WikiNect (2013)
12. Klein, W.: Raumausdrücke. Linguistische Berichte 132, 77–114 (1991)
13. Lücking, A., Bergmann, K., Hahn, F., Kopp, S., Rieser, H.: The Bielefeld speech and gesture alignment corpus (SaGA). In: Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality, LREC 2010, pp. 92–98 (2010)
14. Jörgensen, C., Jaimes, A., Benitez, A., Chang, S.: A conceptual framework and empirical research for classifying visual descriptors. Journal of the Am. Soc. f. Inf. Sci. a. Tech. 52(11), 938–947 (2001)
15. Zipf, G.K.: Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology. Hafner Publishing Company, New York (1972)
16. Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. Cont. Phys. 46, 323–351 (2005)
17. Adamic, L.A.: Zipf, power-law, Pareto — A ranking tutorial (2000), `http://www.hpl.hp.com/research/idl/papers/ranking/`
18. Rohde, H., Clarke, A., Elsner, M.: Wally referring expression corpus (wrec) v0.1.0, (dataset). University of Edinburgh. School of Philosophy, Psychology, & Language Sciences. Linguistics & English Language (2013)