# Defining and Optimizing User Interfaces Information Complexity for AI Methods Application in HCI

Maxim Bakaev(✉) and Tatiana Avdeenko

Novosibirsk State Technical University, Novosibirsk, Russia
bakaev@corp.nstu.ru, avdeenko@fb.nstu.ru

**Abstract.** The HCI has understandably become user-centric, but if we are to consider human operator and computer device as even components of a human-computer system and seek to maximize its overall efficacy with AI methods, we would need to optimize information flows between the two. In the paper, we would like to call to the discussion on defining and measuring the information complexity of modern two-dimensional graphic user interfaces. By analogy with Kolmogorov complexity (algorithmic entropy) for computability resources, the interface information complexity could allow estimating the amount of human processor resources required for dealing with interaction task. The analysis of the current results allows concluding that interface "processing" time by humans is indeed affected by the interface message "length" parameter, and, presumably, by vocabulary size. We hope the results could aid in laying ground for broader AI methods application for HCI in the coming era of ubiquitous Big Interaction.

**Keywords:** Model human processor · Interface design automation · Information complexity · Hick-Hyman's law

## 1 Introduction

Just as the recent emergence of Big Data field became the result of ongoing exponential growth of available and generated data, soon we may face the phenomenon of Big Interaction. The multiplicity and extensiveness of data sources, the diversity of user needs and tasks, as well as of interface devices and contexts of use, may leave us in a situation when hand-making of all the necessary human-computer interfaces by dedicated designers becomes impossible or economically unfeasible. A possible solution is employment of Artificial Intelligence (AI) methods, which may be able to ensure "good enough" interaction.

Indeed, there were already approaches and even products proposed that are able to automatically generate user interfaces for relatively simple tasks or for special contexts of use. For example, the usage of PUC system dedicated to the creation of standardized interfaces for various home appliances was reported to improve the interaction quality rates 2–4 times [1, p. 185]. In it, the language specially developed for describing interface models was simplified based on peculiarities of the task at hand – as such, it didn't allow specification of users tasks, because home appliances by and large don't

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2015, LNCS 9170, pp. 397–405, 2015.
DOI: 10.1007/978-3-319-20916-6_37

imply complex multi-operational interaction. The interface code generation was based, in particular, on heuristic rules providing standardization of the interfaces' visual appearances and on ontological system for describing semantic equivalence of concepts used in various appliances descriptions.

Another example is SUPPLE system, used for creating alternative interfaces for users whose needs weren't considered in mainstream interface of a product or device, and for which the average increase in effectiveness of 62 % was reported [2, p. 45]. The authors noted that the system was most suitable for creating standard interfaces based on dialogue windows, because there is well-established taxonomy for them, which describes the possible interaction elements. The interface generation was considered as a discrete optimization problem, while input data were functional description of interface, the model of the platform's capabilities and limitations, the interface usage model, and criterion function incorporating parameterized quality indexes. This function would mostly cover "physical" parameters in interaction, such as movement time between interface elements, or their size. It should be noted that it was also able to consider the usage of interface by a specific user category, e.g. people with motor disabilities.

The above products could be said to belong with the so-called model-oriented approach, when abstract user interface is specified (PUC or SUPPLE), or somehow derived – from existing programming code (such as in Mickey or HUMANOID), from database model (GENIUS) or from high-level user tasks. The actual interface code generation is then based on knowledge-base rules or on optimization of certain interface parameters or expected quality indexes [2, pp. 3, 4]. Yet another representative here is RAS IACP's system that allows to denote design resolutions and to perform automated interface quality validation [3]. The system was also based on ontological approach that implies specification of concepts related to user interfaces, such as user tasks, use cases, information presentation, etc. When the interface model was automatically transformed into code, the interaction quality was insured via usability metrics also existing in the ontology, together with specialized language for specification their calculation algorithms.

All in all, the review of AI methods applicability in the HCI field could be summarized as the following directions, listed in the order of increasing intellectuality:

- Recommendation of design resolutions or providing relevant guidelines/patterns for user interface being created by human designer. Indeed, the so-called tools for working with guidelines are quite widespread, although the relevance issue remains problematic, which hinders their practical use (see reviews and reflections in [4, 5]).
- Validation of available user interface code and identification of errors or disadvantages [3]. Currently, automated validation tools mostly cover syntactical aspect only, and can hardly understand semantics. Some approaches for deeper analysis are proposed, in particular ones based on domain ontologies, but the involved prior effort generally outweighs the automation benefits, similarly to interface code generation mentioned below.
- Interpretation of available user interface code and the adjustment of ensuing interaction to match user needs, characteristics of interface devices, etc. E.g. modern web-browsers in a relatively non-intelligent way can vary webpage presentation due to many factors, mitigate code errors, etc.

- Designing user interface and generating its code based on provided specification [1, 2]. However, detailed specification of user interface using a formal language or as an interaction model becomes too extensive as the complexity of interaction increases, and the effort required to spell it out quite soon exceeds the one needed for making an actual user interface.
- Creating specification for the user interface based on "understanding" of full interaction context and the functions of the involved software, predicting user needs, etc. [6]. It seems a promising and possibly feasible approach for the Big Interaction era, although the results are likely going to remain somehow close to interface wireframes and be aesthetically inferior compared to "hand-made" solutions.

It should be noted, however, that the widespread optimization-based automated interface code generation has a fundamental problem in the Big Interaction era. Most often neither a designer nor a supporting system would have confidence in how exactly the interface is going to show. Obviously, optimizing distances between interface elements and their sizes (like in SUPPLE) has little sense for a web interface code processed and shaped by a web browser, not even considering varying screen sizes. Thus, we believe, the optimization could be founded on different principles, such as the measurements of information volumes transferred between human and computer. In our paper we call to the discussion on defining and measuring the information complexity of modern two-dimensional graphic user interfaces (GUI), which may be loosely based on Kolmogorov complexity (algorithmic entropy) and Halstead's software metrics. Possibly, information complexity can dictate optimal user interface structure and content, and lay ground for broader AI methods application in HCI.

## 2 Methods

There is no lack of study of "interface devices" present in a human body – for example, human's visual system throughput is estimated at 50–70 bit/s for passive perception of images (e.g. watching television), while for reading that implies comprehension the value drops to at least 30–40 bit/s. For the output tasks, speech allows up to about 50 bit/s, writing with a pen – 10 bit/s, while computer mouse and keyboard are at 3–5 bit/s and up to 25 bit/s respectively [7]. Currently, the applications of these data are quite limited, because there seems to be no accepted way to measure the amounts of information transferred between human and computer via user interfaces, except for simplest cases. The most straightforward way to quantify information contained in a user interface would be application of Hick-Hyman's Law, known to HCI researchers for already quite a long time.

### 2.1 Hick-Hyman's Law in HCI

As selection tasks that are prevalent in many modern interfaces may be represented as combination of choice and movement stages, the application of the infamous Fitts' and Hick's laws for modeling would seem a natural approach. We'd like to remind that W.E. Hick, applying Shannon's Information theory to psychological problems, observed that reaction time (RT) when choosing from N equiprobable alternatives is proportional to the logarithm of their number:

$$RT \sim k * \log_2(N + 1), \tag{1}$$

where k is the rate of gain of information. Later, R. Hyman reasonably noted that RT is in fact linearly related to information quantity, i.e. the entropy of the set of stimulus ($H_T$):

$$RT = a_H + b_H * H_T, \tag{2}$$

where $a_H$ and $b_H$ are empirically defined constants. The slope in thus formulated Hick-Hyman law (2), $b_H$, in simplest cases is believed to be equal to 150 ms, then the corresponding Hick's rate of gain of information ($b_H^{-1}$) is equal to 6.7 bits/s [8].

Unlike the Fitts' law that adequately models movement sub-stages, Hick's law generally falls short to describe cognitive performance, as the amount of information that needs to be processed (HT) is far more complex than log2(N+1) for any real tasks [9, p. 341]. With the experimental investigation described below we sought to improve the information measure and propose alternatives to the Hick-Hyman's law [10], so far by incorporating in the model visual search time, as a measure of information complexity.

### 2.2 The Experimental Investigation

**Subjects.** Twenty eight subjects took part in the experiment. Fifteen participants (4 male, 11 female) were elder people and their age ranged from 56 to 74 (M = 63.4, SD = 5.26), recent graduates of 36-h computer literacy courses held by People's Faculty of Novosibirsk State Technical University (NSTU). Thirteen subjects (5 male, 8 female) were recruited among NSTU students and general staff. They ranged in age from 17 to 30 (M = 23.9, SD = 4.38). All subjects had normal or corrected to normal vision. Eight (53.3 %) elder subjects reported having no experience in using computers or mouse before the computer literacy courses.

**Experiment Design and Procedure.** The experiment consisted of two parts: in the first (control) one the subjects were assigned typical movement tasks modeled with Fitts' law, while in the second one the participants were asked to perform selection tasks. The general experiment design was carried out in accordance with recommendations for Fitts' law experiments, provided in [11]. It was within-subjects, with two groups of participants – elder people and younger computer users. Before the experiment, data regarding the participants' age and gender were gathered. All subjects participated in the experiment voluntarily, and prior to the experimentation informed consents were obtained. Each subject then did a test run of trials with random combinations of A (distance to target), W (targets size), and N (number of targets in the second experiment), until fully understanding the assignment, to negate the effect of practice.

In the first experiment, the two main independent variables were size of a square target (W: 8, 16, 32, 64, 128) and distance to it (A: 64, 128, 256, 512, 1024). There were 7 different ID values (not all combinations were used), ranging from 1.58 to 7.01. The number of outcomes for each combination of A and W was lower than generally recommended (15 for each of ID values), because of the exploratory nature of the study and the intent not to tire the seniors.

The subjects were presented with two squares, a starting position and a target, dissimilar in shape and color. They were positioned randomly in relation to each other on a computer screen to negate the effect of movement direction. The subjects were asked to click the starting position with a mouse pointer and then, "as fast and as accurately as possible", move the pointer to the target and click it. Coordinates of both clicks were recorded; also if the second click was outside the target, error was recorded, and participant was taken to a next trial. The dependent variables were performance time (MT, between the two clicks) and error ($E_1$).

In the second experiment, the target would become visible on the screen only after participant's click on the starting position. False alternatives (of dissimilar shape and color, all of them identical, so overall vocabulary size n = 2) would appear together with the target. The number of alternatives was additional independent variable with 3 levels (N: 2, 4, 8), which were so far deliberately chosen not to exceed Miller's number of 7 ± 2. Also, there were A and W resulting in 6 different values of ID, ranging from 1.58 to 6.02, with 17 outcomes for each level of N. Again, the dependent variables were performance time (ST, between the two clicks) and error ($E_2$, clicks outside the target).

To measure and record the values of independent and dependent variables, an online application was developed with PHP and MySQL and used in IE web browser, with performance time measured with JavaScript to eliminate any server-side delay. The sessions with the two groups of participants, elder and younger, took place with 21-days interval in a same room on same computer equipment, with monitor screen resolution of 1024*768 pixels (thus constant $S_0$ of 1000*600 pixels).

**Hypotheses.** To confirm our reasoning, we identified several hypotheses to be checked in the subsequent experimental investigation:

H1. There is performance difference (time, accuracy) between movement and selection tasks.
H2. Hick's law is not adequate to model selection time.
H3. Visual search time is appropriate addition to movement time in modelling selection tasks.
H4. The proposed model is robust enough to plausibly model performance for different user groups.
H5. Movement and selection throughputs correlate per subjects and are affected by identical factors.

## 3 Results

**First Part (Movement Tasks).** The 15 outcomes for each of 7 ID values in the first part of the experiment resulted in 105 data for each participant, producing a total of 2940 data, of which 2888 (98.2 %) were considered valid. Invalid were the outcomes

**Table 1.** Mean MT and $E_1$ per Fitts' ID

| ID | 1.58 | 2.32 | 3.17 | 4.09 | 5.04 | 6.02 | 7.01 | Mean (SD) |
|---|---|---|---|---|---|---|---|---|
| MT, ms | 468 | 617 | 777 | 890 | 1039 | 1247 | 1425 | 922 |
| | (251) | (303) | (379) | (374) | (395) | (483) | (507) | (503) |
| $E_1$, % | 3.4 | 5.6 | 4.6 | 4.8 | 5.6 | 6.8 | 11.0 | 6.0 |

when subjects made an obviously erroneous click far from target or when the registered time was higher than 3000 ms. Table 1 shows mean values for movement time (MT) and error level ($E_1$) per Fitts' ID as well as overall ones.

MANOVA was used to test the effect of subjects' characteristics such as subject group (elder or younger), gender and experience (for this factor, the analysis was done for elder participants only) on MT and $E_1$. The effect of the experimental conditions in the first experiment was analyzed independently for the two subject groups. Predictably, distance (A) had significant effect on MT for both elder and younger participants. At the same time, the effect of distance was not significant for the number of errors committed by neither seniors ($F_{6,1502}$ = .9; p > .5), nor their younger counterparts ($F_{6,1336}$ = 1.2; p = .29). Size of target (W), besides significantly affecting MT for both subject groups, also had significant effect on error level for both elder ($F_{4,1502}$ = 5.5; p < .001) and younger participants ($F_{4,1336}$ = 2.7; p = .03). Post-hoc analysis indicated that only W = 8 px was significantly different in terms of committed errors, for both groups, and led to 10.2 % and 12.3 % errors for elder and younger subjects respectively. The interaction between distance to target and its size was not significant for either of the subject groups.

**Second Part (Selection Tasks).** The number of outcomes for each participant in the second part of the experiment was 51, producing a total of 1428 data, of which 1408 (98.6 %) were considered valid. Table 2 shows means for selection time (ST) and error level ($E_2$) per ID and number of targets (N) as well as overall ones.

**Table 2.** Mean ST (ms) and $E_2$ (%) per N and Fitts' ID

| ID N | 1.58 | 2.32 | 3.17 | 4.09 | 5.04 | 6.02 | Mean (SD) |
|---|---|---|---|---|---|---|---|
| 2 | 842 | 965 | 1016 | 1064 | 1238 | 1467 | 1034 |
| | (318) | (423) | (409) | (283) | (405) | (530) | (414) |
| | 3.6 % | 7.2 % | 6.4 % | 8.4 % | 7.3 % | 7.1 % | 6.6 % |
| 4 | 814 | 953 | 1016 | 1121 | 1259 | 1660 | 1051 |
| | (338) | (432) | (426) | (379) | (399) | (558) | (459) |
| | 4.8 % | 2.8 % | 3.7 % | 9.5 % | 7.1 % | 14.3 % | 5.7 % |
| 8 | 797 | 977 | 1020 | 1170 | 1328 | 1526 | 1061 |
| | (315) | (480) | (392) | (424) | (478) | (439) | (461) |
| | 4.8 % | 6.4 % | 9.1 % | 8.4 % | 7.1 % | 24.0 % | 8.1 % |
| Mean (SD) | 818 | 965 | 1018 | 1118 | 1275 | 1552 | 1049 |
| | (323) | (444) | (408) | (368) | (428) | (514) | (444) |
| | 4.4 % | 5.5 % | 6.4 % | 8.8 % | 7.2 % | 14.8 % | 6.8 % |

As in the first part of the experiment, a multivariate analysis of variance was used to test the effect of subject group and gender on ST and $E_2$. The results suggest highly significant effect of subject group on time ($F_{1,1404}$ = 365.8; p < .001), with estimated marginal means of 1238 ms for elder subjects vs. 814 ms for younger ones. The effect of subject group on error was not significant ($F_{1,1404}$ = 1.8; p = .18), in contrast to the first part of the experiment. The gender factor remained significant for both ST ($F_{1,1404}$ = 5.3; p = .022) and number of committed errors ($F_{1,1404}$ = 5.0; p = .026). As before, male participants on average were somehow faster, with 1001 ms vs. 1051 ms for female ones. The mean number of errors was 4.6 % and 7.8 % respectively. No significant interaction between the independent variables was observed.

**Visual Search Time.** To further examine the effect of N on ST (which did not clearly manifest in Table 2), we ran MANOVA test with ST and $E_2$ as dependent variables and N, W and A as factors. We found no significant effect for N on neither ST ($F_{2,1357}$ = .27; p = .76), nor $E_2$ ($F_{2,1357}$ = 1.03; p = .36). The effect of W was highly significant for both ST ($F_{3,1357}$ = 131.36; p < .001) and $E_2$ ($F_{3,1357}$ = 6.05; p < .001). Movement amplitude A significantly affected ST ($F_{3,1357}$ = 23.51; p < .001), but not $E_2$ ($F_{4,1357}$ = 2.04; p = .09).

We attempted preliminary regression models for ST with $\log_2$(N) and Fitts' effective index of difficulty as factors, and N was not significant in the regression (p = .308), so we decided to exclude the number of objects from the visual search time model. Thus, we proposed the index of visual search difficulty ($ID_{VS}$) in the following form:

$$ID_{VS} = \log_2(S_0/S) = \log_2(S_0/W^2), \qquad (3)$$

where S is equal to $W^2$ in case of our square targets. The justification is twofold:

1. The $S_0$/S represents the "length" of graphic interface as a message, i.e. the maximum number of elements of square S that it can contain. It seems reasonable to assume that users "process" not just the displayed objects, but the whole interface, including whitespace. Then $S_0$/S should take the place of N in Hick's law (1).
2. Parallels may be also drawn with motor behaviour described by Fitts' ID: then "search amplitude" $S_0$ corresponds to A and "search termination area" S – to W.

**Hypotheses Check Results.** Based on the analysis described in more detail in [10], we can make the following conclusions regarding the previously stated hypotheses.

H1. Confirmed. Selection tasks took more time to complete and the accuracy was lower than for movement tasks. We'd like to note that the increase in performance time was nearly constant for the two subject groups, 187 ms for elder vs. 236 ms for younger participants, but the growth in error level for senior subjects was far more dramatic, at +75.0 %, which may be explained by poorer multi-tasking abilities of people in older age.

H2. Confirmed. The number of alternatives (N) didn't have significant effect on ST, and $\log_2(N+1)$ was not significant in the regression.

H3. Partially confirmed. $ID_{VS}$ (3) was significant in regression for ST, but the resulting $R^2$ were lower than for movement tasks (see in [10]).

H4. Confirmed. ST regressions with $ID_{VS}$ factor were significant for both elder and younger subjects, and regression coefficients suggest that visual search task is relatively harder for seniors than movement task. This corresponds well to sharp increase (+75 %) in error level for elder participants in selection tasks.

H5. Confirmed. Movement and selection throughputs are relatively highly correlated per subjects, and the effects of age and experience Fitts' throughput (TP) and TPS (see its formulation in [10]) are similar.

## 4 Conclusions

In our paper we raised the problem of quantifying information flows in human-computer systems, via introducing information complexity measure for user interfaces. The straightforward information entropy approach (2) has proved to be problematic in real circumstances, so we proposed to use visual search difficulty (3) to reflect the graphic user interface complexity. Search area size ($S_0$), sought element size (S) and the number of alternatives (N) were elected as primary factors for VST, while also employing vocabulary size parameter is the goal of our next experimentation. In the result of experimentation with 28 subjects of different age groups (described in more detail in [10]), visual search difficulty was suggested as the logarithm of the ratio between $S_0$ and S, with N not being significant.

Thus obtained mean value for proposed selection task throughput (TPS), 12.6 bit/s, seems to be consistent with established human visual processing capacity that ranges from 5 to 70 bit/s. It is known that tasks requiring deeper processing have lower capacity: perception of TV picture is at 50–70 bit/s, simple text reading – 40–50 bits/s, while text proof-reading – 18 bit/s [7, p. 62], so TPS was to be expected in the lower part of the range. However, the model is subject for further development, and we expect that the information complexity measure should be $ID_{VS}$ multiplied by the vocabulary size – how many kinds of different objects, i.e. interface elements, are employed on the screen. Further development of our research implies closer analysis of the classic Kolmogorov's algorithmic entropy and Halstead's software "difficulty" measures.

We believe that $ID_{VS}$ or the enhanced information complexity measure could be used in optimization when auto-generating user interfaces, as they are independent of absolute size measurement, which is of particular importance in adaptable web interfaces or multitudinous mobile interfaces. As we noted before, greater degree of AI methods utilization for creating user interfaces may be deemed necessary to cope with the Big Interaction, caused by ever-increasing diversity of users, their tasks, interface devices and contexts of use.

## References

1. Nichols, J., Myers, B.A.: Automatically generating high-quality user interfaces for appliances. Doctoral dissertation, Carnegie Mellon University, Pittsburgh (2006)
2. Gajos, K.Z., Weld, D.S., Wobbrock, J.O.: Automatically generating personalized user interfaces with Supple. J. Artif. Intell. **174**(12–13), 910–950 (2010)
3. Gribova, V.V.: Automation of design, implementation and maintenance of user interface based on ontological approach. Doctoral dissertation, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Science (2007) (in Russian)
4. Dearden, A., Finlay, J.: Pattern languages in HCI: a critical review. Hum. Comput. Interact. **21**(1), 49–102 (2006)
5. Chevalier, A., Fouquereau, N., Vanderdonckt, J.: The influence of a knowledge-based system on designers' cognitive activities: a study involving professional web designers. Behav. Inf. Technol. **28**(1), 45–62 (2009)
6. Bakaev, M., Avdeenko, T.: Indexing and comparison of multi-dimensional entities in a recommender system based on ontological approach. Computación y Sistemas **17**(1), 5–13 (2013)
7. Gasov, V.M., Solomonov, L.A.: Engineering psychology design of human interaction with technical instruments vol. 1. Visshaya shkola, Moscow (1990) (in Russian)
8. Longstreth, L.E., et al.: Exceptions to Hick's law: explorations with a response duration measure. J. Exp. Psychol. Gen. **114**, 417–434 (1985)
9. Seow, S.: Information theoretic models of HCI: a comparison of the Hick-Hyman law and Fitts' law. J. Hum. Comput. Inter. **20**(3), 315–352 (2005)
10. Bakaev, M., Avdeenko, T., Cheng, H.I.: Modelling selection tasks and assessing performance in web interaction. IADIS Int. J. Comput. Sci. Inf. Syst. **V VII**(1), 94–105 (2012). Isaías, P., Paprzycki, M. (eds.)
11. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research. Int. J. Hum. Comput. Stud. **61**(6), 751–789 (2004)