What Do Users Really Like in Menus: Building Menu Optimization Criterion

Mikhail Goubko
V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences
65 Profsoyuznaya str.
Moscow, 117997 Russia
mgoubko@mail.ru

Alexander Varnavsky
Ryazan State Radio Engineering University
59/1 Gagarina str. Ryazan,
390005 Russia
varnavsky_alex@rambler.ru

Abstract
In recent years several computer-aided design (CAD) tool prototypes were developed for the design of hierarchical user menus. These CAD systems differ in optimization criteria used to compare menus. A traditional approach relates menu performance to the target item selection time, but is the faster always the better for a user? We elicit key factors determining the user valuation of a menu-driven search from the laboratory experiment and build the empirically grounded function of user satisfaction, which depends on the selection time and accuracy, but also on menu logical compliance.

Author Keywords
menu-driven system; hierarchical menu optimization; optimization criterion identification; menu testing and assessment

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): User Interfaces.

Introduction
Since 1980s, methods for the automated design of hierarchical user menus were a valuable application point of formal methods to human-computer interaction. On the one hand, the interest of researchers was inspired by the wide use of hierarchical menus in user interfaces

Note:

A yet another recognizable metric is the user selection **error rate**. It accounts for the extra displeasure of users selecting a wrong menu item. Yet, the error rate is rarely predicted in the literature.

Note:

The alternative to selection time as a menu performance metric is direct user satisfaction assessment via filling questionnaires (see [3], [5]). The open issue here is prediction of partial valuations from the menu design and structure.

(e.g., in the standard GUI). On the other hand, the process of user navigation in menus appeared a convenient subject for the formalization.

A CAD system for user menu design automates the construction of a menu and helps building better menus by suggesting efficient and clear menu structures. The key aspect of menu design automation is the choice of a relevant menu optimization criterion. Recently several approaches to menu structure optimization were reported [1], [7], [8], [9], [11], [15] employing various combinatory algorithms to optimize different metrics of menu performance.

Traditionally, the target item selection time was considered a key performance metric of a user menu (e.g., see [13]). The idea is that menu navigation is always a reluctant (non-productive) activity, and users want to get rid of it. To calculate the selection time during the menu design process one needs a predictive model based on menu structure and design. Existing models [2], [4], [6], [10], [12], [14] account for the menu breadth, labels' size and popularity, alphabetic sorting, but neither for the semantics of labels (such as label length or clarity) nor for the logical or cognitive consistency of labels grouped together on a single panel.

Thus, correction terms are added to the criteria in [1], [11], [15] to care for labels' consistency. Consistency is an optimization constraint in [8] and [9], while in [7] the *semantic quality* of a menu item affected its perception time. The structure of the criterion is rarely discussed in this literature, and it is still an open question whether correction terms just enhance accuracy of selection time prediction models, or they reflect an independent aspect of the menu valuation by a user. In the other words, whether users really prefer menus where they navigate faster to the slower ones?

We justify the set of factors relevant to user satisfaction of a menu and reveal the relevant structure of a criterion for menu optimization algorithms. In our experimental setup subjects assess different hierarchical menu structures after performing a series of searches. Then we relate satisfaction scores to metrics of the menu structure and to those of the navigation process.

Menu score factors can be divided into five groups (see Table 1). The first four groups are independent, while process variables can be predicted. We approve that the empirical selection time essentially depends on menu consistency, and, therefore, menu consistency must be incorporated in future time prediction models to increase their accuracy. Then we relate users' menu valuation to the structure and process variables and show that selection time is not enough to carefully predict menu scores. The users exceptionally reward clear menu logic even if item selection time is not decreased.

Experimental Design

Volunteers (48 graduate students of the Engineering department of Ryazan Radio Engineering University) assess hierarchical menus solving standard tasks of finding a metal-working tool with a sequence of choices of its features on menu panels. The features include the tool name (drill, mill, cutter), manufacturer (YG-1, Dijet, QCT, Hammond, HAM, Carmex), length (shorten, normal, long), the diameter (micro, mini, small, normal, big, extra), the use (rough cut, finishing, uncooled cut), the material (carbide, high-speed steel, replaceable insert). A task definition presented to users combines text, iconic, and synonymic stimuli (Figure 1).

1 Menu design variables describe visual design of a menu (the menu is linear, iconic, or split; the font size, colors, and items' layout)

2 Menu structure variables characterize the menu hierarchy (menu breadth, depth, labels' consistency, etc)

3 User characteristics fix the character, computer skills, cognitive, or motor abilities of a subject.

4 Task variables explain the usage context (time pressure, repetitiveness of tasks, learning time, price of an error).

5 Process variables characterize the process of menu navigation – item selection time, error rate, number of menu cancellations, etc.

Table 1. Groups of variables affecting menu valuation

We use a pretty complex (yet, realistic) experimental setup based on the standard Delphi GUI. Ten distinct menu structures are built by distributing tool features to different menu levels (see Figure 2). Some panels may join two or more features (e.g., the menu item is "Carmex, finishing".) Menus 1-6 are *symmetric* (i.e., the order of feature selection is the same for all target items.) Menus 7-10 are *skewed* (e.g., Menu 7 inherits the structure of Menu 1 for drills and cutters and the structure of Menu 5 for mills.) All menus share the same set of 324 target items. Each subject tests sequentially eleven menus (Menu 11 is a copy of Menu 1, which is not announced), having 5 minutes for each menu and then assigning an overall score from -5 to 5 (each grade is equipped with a text description.)

A task picture is displayed for 30 sec, and then the next of 12 tasks is displayed in a cycle. User actions (mouse and keyboard) are logged each 10 msec. Before the experiment the subjects passed psychophysiological tests (Eysenck personality, Ravena IQ, Landolt rings, and numbers memorizing) for future use.

Analysis

We use a linear regression to predict normalized score from the structure and process variables. All trials of a user in a menu are averaged giving the total of 528 cases. A number of derivative variables based on time and accuracy are calculated: task execution period T, successful task completion time TS, task success rate NS, average time of a user AT, time to average T/AT, time normalized to success rate $T/(AT \cdot NS)$, execution time and success time for the first 12 trials (to measure menu adoption speed). The menu structure is described with the menu depth and breadth. Logical consistency of a menu is captured with the number of "joined" panels J (containing items like "Carmex, finishing"), skewed menu dummy, and the variables caring for the order of feature selection. User scores are normalized to balance optimists and pessimists.

We also predict selection time from the menu structure and logic with nonlinear models similar to [2] and [6] both for the original data (7846 cases) and for user-averaged data (132 cases.)

Results

Theoretically, navigation in a hierarchical menu reduces to a sequence of submenu selections. The time spent in a submenu adds up from stimulus perception, item search and mouse movement (probably, in parallel [2]), and, finally opening a submenu. Popular predictive models (like [2] or [6]) do not account for stimuli' and items' heterogeneity, thus, poorly predicting navigation time in our setting. Individual user behavior seems too random to explain. For user-averaged data the best prediction for the errorless navigation time (corrected R^2 = 0,852) uses empiric weights of distinct menu panels multiplied to menu, menu panel, and menu item exponential learning terms. Surprisingly, menu geometry (breadth, item number, mouse movement amplitude, etc.) insignificantly affects the average time. We conclude that the menu logic drives user selection behavior rather than menu geometry in our setting.

Cluster analysis of score vectors gives a single cluster. Therefore, user personality can be neglected and we can concentrate on the structure and process variables. The correlation of the normalized score with process variables is 0.567 (corr. R^2 = 0.304), with structure variables – 0.536 (corr. R^2 = 0.277), while their combination gives correlation 0.696 with corr. R^2 = 0.467.

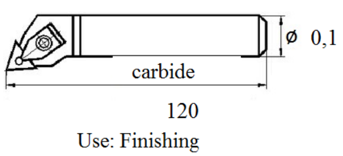

Figure 1. A task definition

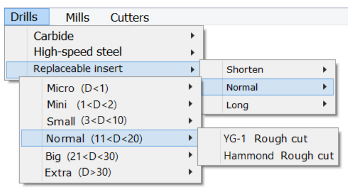

Figure 2. The Menu 1 (translated from Russian)

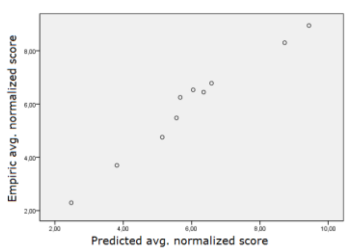

Figure 3. Scattering diagram: Avg. score vs (avg. *TS* + *J* + "Top panel is Type")

Both empiric time and accuracy are significant process variables, while skewed and combined menu count, and the order of feature selection are significant structure variables (students noticed informally they find natural a certain order of feature announcement, e.g. instrument type or a manufacturer first.) For the user-averaged data the score is predicted with corr. R^2= 0,983 using success time *ST*, No of joined panels *J* and "Top panel is Type" dummy (see Figure 3.) The conclusion is that user preferences go beyond selection time, and optimization criterion should care explicitly for the menu consistency, in particular, for the natural item groupings and the order of groups' appearance.

Therefore, since only figures averaged over a user population can be accurately predicted, a menu optimization criterion **has to be** a linear combination of predicted navigation time, user error rate, and some function of menu logic variables. The predictive models should account for cognitive characteristics of menu panels and individual menu items (e.g. item perception time).

Acknowledgements

We thank the volunteers participated in menu testing. Author 1 gratefully acknowledges the RFBR grant 13-07-00389 and Author 2 – the RFBR grant 14-38-50904.

References

[1] Bailly, G., Oulasvirta, A., Koetzing, T., Hoppe, S. MenuOptimizer: Interactive optimization of menu systems. *Proc. UIST 2013* (2013), 331-342.

[2] Bailly, G., Oulasvirta, A., Brumby, D.P., Howes, A. Model of Visual Search and Selection Time in Linear Menus. Proc. CHI 2014, ACM Press (2014),

[3] Brooke, J. SUS: A Retrospective. *J. of Usability Studies 8*, 2 (2013), 29-40.

[4] Callahan, J., Hopkins, D. Weiser, M., Shneiderman, B. An empirical comparison of pie vs. linear menus. *Proc. CHI 1988*, ACM Press (1988), 95-100.

[5] Chin, J.P., Diehl, V.A., Norman, K.L. Development of an instrument measuring user satisfaction of the human-computer interface. *Proc. CHI 1988*, ACM Press (1988), 213-218.

[6] Cockburn, A., Gutwin, C., Greenberg, S.A. Predictive model of menu performance. *Proc. CHI 2007*, ACM Press (2007), 627-636.

[7] Danilenko, A.I., Goubko, M.V. Semantic-aware optimization of user interface menus. *Automation and Remote Control 74*, 8 (2013), 1399-1411.

[8] Fisher, D.L., Yungkurth, E.J., Moss, S.M. Optimal menu hierarchy design: syntax and semantics. *Human Factors 32*, 6. (1990) 665-683.

[9] Goubko, M.V., Danilenko, A.I. An automated routine for menu structure optimization. *Proc. of 2nd EICS*, ACM Press (2010), 67-76.

[10] Lee, E., MacGregor, J. Minimizing user search time in menu retrieval systems, *Human Factors* 27 (1985), 157-162.

[11] Matsui, S., Yamada, S. Genetic algorithm can optimize hierarchical menus. *Proc. CHI 2008*, ACM Press (2008), 1385-1388.

[12] Norman, K.L. *User response time for pulldown menus as a function of knowledge of location and structure of menu*. Working Paper, Univ. of Maryland, 1988.

[13] Norman, K.L. *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. Ablex Publishing Corporation, 1991.

[14] Paap, K.R., Roske-Hofstrand, R.J. The optimal number of menu options per panel. *Human Factors 28*, (1986), 377-385.

[15] Troiano, L., Birtolo, C., Armenise, R., Cirillo, G. Optimization of menu layouts by means of genetic algorithms. In *Evolutionary Comput. in Combinat. Optimization*, Springer, Berlin-Heidelberg (2008), 242-253.