Human Systems Engineering and Design (IHSED2025), Vol. 198, 2025, 48–52
https://doi.org/10.54941/ahfe1006760

Research on Quantitative Assessment Method of Operational Performance Based on GOMS Cognitive Modeling

Kaili Yin[1,2], Zhanshuo Zhang[1,2], and Ziang Chen[1,2]

[1]National Key Laboratory of Human Factors Engineering, Beijing, 100094, China
[2]China Institute of Marine Technology & Economy, Beijing, 100081, China

ABSTRACT

User operation performance is the most important index to reflect the effect of human-computer interaction design of information system, and the evaluation result of user operation performance can directly reflect the design effect of information system, but there is no quantitative evaluation method of operation performance at present. In this study, the user operation performance evaluation of information system is oriented, and by carrying out ergonomics experiments based on trackball operation mode, the meta-operation time in the GOMS model is corrected and supplemented, and the quantitative evaluation method of operation performance is researched based on the corrected results of the GOMS model. The results show that the method can obtain the quantitative assessment results of user operation performance, which can support the human-computer interaction design enhancement of the information system.

Keywords: Human-computer interaction design, Operational performance, Quantitative evaluation

INTRODUCTION

The common indicators of user operation performance include correct rate and operation time, in which the correct rate is a quantitative result, which can get a more intuitive quantitative assessment results, but the operation time is a time- based data, which can be used to compare and analyze the operation time of the unified task in different design schemes of the information system, but when faced with the single design of the information system for the assessment of the performance, it is not possible to get quantitative assessment results (Xiang, 2011).

The GOMS model, as a cognitive model that predicts on its own the sequence of behaviors and the time required to complete them, helps designers and engineers to accurately describe and analyze the user's process of using the system, and thus to address the relative effectiveness of different scenarios in terms of their timeliness (Xiang, 2011). Therefore, by using the GOMS model to import the user's process of achieving the end goal, the time spent by the user interacting with the interface can be tested, and the time taken to perform each step of each task can be counted and compared with the user's homework time (Wang, 2021).

The decomposition of GOMS behaviors and goals suggests that it can be applied not only to qualitative studies of interactive interfaces but also to quantitative analyses (Xiang, 2011). The amount of time a user spends to accomplish a goal is the total amount of time spent by the user interacting with the interface for each of the basic operations. From the perspective of quantitative analysis, it can predict the time required for users to use a functional interface, evaluate the performance of the interface, and improve and optimize the existing interface (Jiaying, 2023).

GOMS Model Correction Experiments

The GOMS basic operation schedule contains most of the meta-operation time after the decomposition of the basic operation, but there are some more important meta-operations missing in the typical mission operation environment or meta- operations that are not applicable to the ship operator's operation tools, for example, the need to use the trackball to operate in specific environments, so it is necessary to carry out some corrections and supplementary experiments on the meta- operation time in the GOMS model (Jones, 2015).

To design the trackball movement experiment, first verify whether the trackball movement conforms to Fitt's law, and explore the empirical parameters a and b of the trackball mouse using linear regression.

The participant moves the trackball mouse to the starting point and clicks once, then starts recording the time. After moving the mouse to any position within the target area and clicking again, one time measurement is completed.

Designed Experiments

(1) Fitt's law verification experiment for trackball movement time: The distance from the starting point to the target area is divided into three levels in the reference literature: 96mm, 144cm, and 192mm. The target area size is specified as 3, 6, and 12mm. The experimental combinations of d and s need to be randomized. Each combination of starting point and target area appears randomly on the screen three times. Record the t, D, and S for each experiment (Caesaron, 2022; Bruce, 2020).

(2) Experiment on the relationship between mouse cursor properties and time: Level 2 of line segment type: straight line, arc line (the subjects were informed of the approximate shape of the arc line in advance); Level 2 of drawing direction: starting point on the left side of the target point, starting point on the right side of the target point; randomly generate the starting position and target point (but distinguish whether the starting position is to the left or right of the target point). Each level is repeated 5 times, with 2x2x5=20 trials per person. The experiment records the distance between the two targets, drawing time, drawing direction, and type of line segment drawn (Sternberg, 2016; Albert, 2013).

Analysis of Experimental Results

(1) Trackball movement time correction experiment: After organizing the experimental data from 10*3*3*3=270 groups (10 subjects * 3 levels * 3 levels * 3 repetitions), linear regression analysis was conducted given the known linear relationship between variables. Additionally, hypothesis analyses such as model residual independence tests were incorporated to validate the regression analysis.

In the analysis of variance, the F-test yielded a significance value of 0.034, which is less than 0.05. This indicates that the linear regression model established between "time taken to move the mouse" and "log2 (D/S + 1)" has significant statistical significance. Specifically, an increase in log2 (D/S + 1) corresponds to an increase in the time taken to move the mouse, which aligns with Fitt's law.

Table 1: Variance analysis of linear regression model.

Serial Number	Model	Variance Test F-Value	Significant P-Value
One	Regression analysis	4.526	0.034

Next, we need to verify whether the above data is suitable for regression analysis, so we will conduct a residual analysis.

Based on the relationship between the regression coefficients, we conclude that in the null hypothesis regression equation, b = 0.148 and a = 2.414.

(2) Mouse stroke attribute and time relationship model: For the dataset with 10*2*2*3*2=120 (10 subjects * 2 line types * 2 directions * 3 line lengths * 2 repeated experiments), a multi-factor repeated measures analysis of variance (ANOVA) in the general linear model was first used to analyze the correlation between the independent variables. Since there were no interactions between factors in the analysis results, the following analysis only describes the influence of individual factors.

The analysis yielded a significant P-value of 0.004, which is less than 0.05, indicating that the data at each time point are statistically significant (i.e., the following analysis is meaningful). The significant P-values for the remaining factors, namely time * target distance, time * drawing direction, and time * line type, are all greater than 0.05, indicating that the effect of drawing time (first measurement, second measurement) obtained from the two repeated experiments does not vary with changes in line type, drawing direction, or distance length.

The significance P-values for target distance and line drawing direction were both greater than 0.05, while the significance P-value for line type was 0.000, which is less than 0.01 but greater than 0.05. This indicates that there is a significant difference between line type and the time taken for line drawing, meaning that different line types affect the time taken for line drawing, while differences in target distance and line drawing direction do not affect the time taken for line drawing.

Therefore, we conducted data analysis based on two types of line shapes. When the drawn line shape is a straight line, the average time taken is 4.76s, and 90% of the time is between 3.16–5.10s, which is similar to the 4s data for drawing a line in the literature. Therefore, it is believed that the approximate time taken to draw a straight line on a screen with this resolution is 4.76s; when the drawn line shape is a curve, the average time taken is 6.34s, and 90% of the time is between 5.03–7.97s. Therefore, it is believed that the approximate time taken to draw a curve using a trackball on this type of screen is 6.34s.

Performance Evaluation Model Based on GOMS Revised Time

The GOMS model provides a time model that can be calculated and measured. It can obtain the meta-operation time of a known operation task and calculate the time predicted by the model, which we refer to as the standard completion time, denoted as t0.

The standard completion time t0 excludes the interference of factors such as emotions, society, and physical environment when people perform actions, representing the ideal operation time under the most harmonious human-machine matching for the tested system carrying the task. Therefore, the quantitative score corresponding to this time is defined as 100.

According to research on user experience metrics, twice the standard operation time is considered a tolerable operation time. We refer to this as the acceptable time, denoted as t1, and the corresponding score is the passing score, denoted as F1. This allows us to obtain the performance evaluation model.

CONCLUSION

This thesis investigates a GOMS cognitive model-based performance assessment methodology, which realizes a quantitative assessment of the operating time of a specific information system task to be assessed, and provides an assessment tool for the training of information system users and the design of human-computer interaction.

However, this experiment was carried out under the resolution of 1920*1080 laptop, and the screen has a certain gap compared to the control screen of the naval ship, so it has some limitations. For example, there is no significant difference between the distance between the start and end points of the drawing line and the time used in the data analysis may be because the screen size is too small and the distance between the two points of the experiment is too short, so the difference in the time used is not significant.

REFERENCES

Albert, B., & Tullis, T. (2013). Measuring the User Experience (Interactive Technologies): Collecting, Analyzing, and Presenting Usability Metrics. Morgan Kaufmann Publishers Inc.

Caesaron, D., Lin, C. J., & Febiyani, M. A. (2022). Evaluating direct pointing and indirect cursor interactions with fitts' law in stereoscopic environments. Journal of engineering and technological sciences, 54(1), 1–17.

Goldstein, E. Bruce. (2020) Cognitive Psychology: Mind, Research, and Life China Light Industry Press.

Li Jiaying, Huo Jiadao, Li Dong, & Huang Ke (2023). Efficiency evaluation of speech gesture interaction based on Goms model Command, Control, and Simulation, 45 (2), 7.

Sternberg, & R., J. (2016). Introduction to optimizing learning in college: Tips from cognitive psychology. Perspectives on Psychological Science, 11(5), 651–651.

Wang Kunlong, Li Changjing, Liu Bing, Song Yingyi, Tian Zongkai, & Miao Xu (2021). A quantitative evaluation method for software usability based on GOMS model improvement CN202110841584.5.

Xiang, L. (2011). On man-machine interface operation performance of vehicles based on bhr-goms model. Design research.