Approach to evaluate effectiveness of human-computer interaction with contemporary GUI

Kostiuk D.A., Derechennik S.S., Shitikov A.V., Latiy O.O.

Brest State Technical University
`d.k@list.ru`

A system to test multi-window interface efficiency is proposed targeted at evaluating interaction changes introduced in contemporary GUI shells. System includes heart rate monitor and EEG scanner. Set of evaluated parameters, registered by the developed software while providing user with typical tasks to execute, includes speed of execution, error tolerance and intensity of the operator's work.

Until recent years GUI of all mainstream operating systems where based on WIMP (window-icon-mouse-pointer) interaction model and desktop metaphor (DM). Screen was interpreted as an analogy of the office table surface, with some objects placed on it: folders and documents, which can be opened ("unfolded" into a window), and some additional utilities [1]. Starting from the year 2011 changes targeted to bring post-WIMP and post-DM features to production are seen in newer versions of GUI shells – in bounds of intension to develop the interface, convenient both for mouse and touch control. Driven by growing popularity of relatively small tablet devices, GUI are gaining alternative application launchers, avoidance of the taskbar and bias to fullscreen-mode applications. The effectiveness of these changes is still discussed, and therefore tools and approaches providing clear picture of human-computer interaction efficiency in modern GUI shells are needed. While carrying out some research in this direction, concerning desktop environments in GNU/Linux [2], we have created specialized set of testing instruments, shown on the following figure:
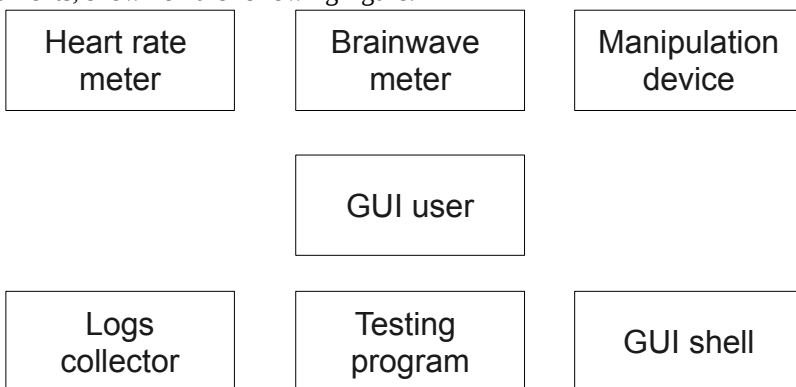
Testing program interacts with GUI shell, providing several windows to work with. Fixed interval of time is set to carry out maximal amount of specified actions with the manipulator (mouse, trackball, touchpad, etc.) or

fixed amount of actions is to be carried out while measuring actual time duration. For now actions to be performed by user are of two types. First type is copying numbers via clipboard, and the second one is visual recognition of geometry figures.

System stores all user's errors in a log file, and user's stage is registered with heartbeat meter (HM) and brainwave monitor (BM), allowing to analyze not only speed and accuracy of operator's work, but also physical and mental load. For now we are using Pulse Sonic K920D and NeuroSky Mindwave as HM and BM respectively.

Testing can be carried out in several stages. At first and most simple stage user has two windows titled as «Source» and «Destination». Window size prevents non-overlapped expose if needed – just to force user actively switching window focus with GUI shell. «Source» window has text field and «Next» button, which sends random number into the text field while being pressed. While one test iteration user is supposed to press the «Next» button, copy number to clipboard with the popup menu, switch focus to «Destination» window and insert number into its text field, also with use of a popup menu (keyboard is not used in this test, if the opposite is not demanded by the experiment requirements). False «Next» button is also present in the «Destination» window to require more concentration. Windows are created by different executable modules to evaluate effectiveness of windows switching for two different applications.

Second stage uses k numbered «Destination» windows and one «Source» window which generates not only a random number to copy, but also shows $x \in [1; k]$ number with a large font, which is the number of the «Destination» window for pasting. We have used $k=3$ in most experiments. Set of actions for one test iteration is the same, but several windows are used to reveal the effectiveness of manipulating the multi-window applications.

Third stage optionally involves «Source» and «Destination» windows with graphical buttons containing randomly shown geometry figures. One button in the «Source» window sets figure to be found in the matrix of 5x5 figures shown in «Destination» window. This experiment copies methodology of the figures memorizing and recognition research, carried out by R.M. Granovskaya and I.Ya. Bereznaya [3].

Information overload is one of the reasons, causing user's work reliability. Defense mechanisms resisting information overload can cause, beside slower reaction, to information skip, its erroneous processing without followed regulation, and prioritized information omission. To register this variety of reactions, four types of errors are stored in a log file: wrong button press, number skip and doubling, and wrong window choice [2].

Four factor groups, considered as a test result, are speed of executing specified actions, number and type of user's errors, mental activity and heart rate.

Speed factor is evaluated via *tempo of test passing* τ, equal to number of test iterations, executed by user in one minute. Taking into account chosen test

stage duration of 5 minutes, τ would be *N/5*, where *N* is number of iterations executed by user while one testing stage.

To provide errors data abstraction from the *N* value for specific user and specific GUI shell, we considered not total number of errors, but medium number of failures *n* per one iteration, and medium length of failure *m*, which is the count of iterations, touched with one failure.

Source parameters to measure heart rate where medium, maximal and minimal values stored by HM for the time of a test stage. Impact of window switching on the workload is obviously lover than the impact of numbers copying or figures recognition actions. Faster method of window switching increases not only number of iterations but also user's tiredness, which in its turn effects the number of iterations in the opposite way.

Brainwaves, registered to evaluate user's mental activity, where captured with the EEG monitor, connected to the open source Puzzlebox Synapse software. The software registered eight types of raw EEG waves, and eSense Brainwave Patterns, calculated by the NeursoSky controller. Two of these eSense values, named as "attention" and "meditation" (correlated to concentration an relaxation degree) where used to compare user's mental activity in different GUI shells.

To decrease impact of the test speed and evaluate how differs the GUI impact at different testing conditions, we compared not heart rate itself, but the integral factor, which was the *intensity of test execution v*, equal to average amount of heart beats in one testing stage, for one iteration in a specific test. This decreased impact of different tiredness values of a specific user at different speed of testing. There is an "effort-tiredness" dichotomy, when optimal zone is specified in the efforts intensity range, surrounded by zones of underruning and overwork with lover labor effectiveness [4]. In some cases weak clutter may increase user's efficiency (in comparison with its absence), and interaction with the elements of a GUI shell, secondary to the main task, may be considered as such clutter. When intensity fits to an optimal zone, this secondary interaction should play role of micro-pauses, i.e. gaps between separate iterations, providing possibilities for short-time rest and preventing the development of tiredness.

References

1) Raskin J. The Humane Interface: New Directions for Designing Interactive Systems. / Addison-Wesley Professional: 2000. – 256 p.

2) Костюк Д., Дереченник С., Шитиков А. Оценка эффективности управления окнами в современных графических оболочках // 7-я конф. «Свободное программное обеспечение высшей школе»: Тезисы докладов / Переславль, 28–29.01.2012. М.: Альт Линукс, 2012. – С. 20–23.

3) Грановская Р.М., Березная И.Я. Запоминание и узнавание фигур. / Л.: Изд-во Ленинградского ун-та., 1974. – 96 с.

4) Забродин Ю.М. Методологические проблемы функционального состояния человека-оператора //Вопросы кибернетики. Психические состояния и эффективность деятельности. М.: 1983. с.3-25.