OPTIONS FOR AUTOMATIC IDENTIFICATION OF USER ACTIVITIES IN USABILITY TESTING

Jan Masner, Jiří Vaněk, Jan Pavlík, Eva Kánská, Edita Šilerová

Department of Information Technologies, Czech University of Life Sciences Prague, Czech Republic

masner@pef.czu.cz, vanek@pef.czu.cz, pavlikjan@pef.czu.cz, kanska@pef.czu.cz, silerova@pef.czu.cz

ABSTRACT

When testing usability of applications, it is often needed to analyze behavior of users in terms of identifying their activities. The activity may be that the user is working on the assignment without problems, is searching for something, is absolutely lost in user interface, is filling a form, is studying the manual, etc. Identifying of the activities is usually done by tagging a video and audio record of the testing, optionally together with visualization of eye movements (eye tracking). It is a very time-consuming work for the usability experts. When testing in a specialized laboratory, we can obtain data from various measurements. Besides audio-visual record, data from eye-tracking, click tracking and keyboard tracking can be analyzed. Moreover, we can engage biometrical data such as pulse, skin temperature, humidity, hand movements, etc. The research question of the paper is whether it is possible to analyze all the data and develop methods and algorithms to automatically identify the user activities. We have performed several experiments in the specialized laboratory for usability testing. The paper describes the research design, experiments, methods of data processing and analysis. Finally, first conclusions and findings are discussed, as well as forthcoming research. The proposed research can be generalized to usability of any products. For example, using glasses for eye tracking and smartwatches, it is possible to conduct usability research of agricultural machinery.

Keywords*:* usability, UX, software, video-tagging, human-computer interaction

1. INTRODUCTION

In general, usability can be defined as the ease of use and learnability of a human-made object. In software engineering, the usability is defined by ISO 9241-11 as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (IDF, 2019). Usability plays an important role in selling any products, whether it is software or equipment for precision farming (Jarolimek et al., 2019). In computer sciences, usability is often extended by using a broader term User Experience (UX) (Benda et al., 2017).

Software products are becoming increasingly complex and it is necessary to consult their development with target group. Therefore, testing usability with end-users has become very important. The testing which utilizes end-users as test subjects can be divided into two basic techniques, testing remotely and laboratory testing (Rubin and Chisnell, 2008). Distance testing requires that the users first start a support application that records their actions (Atterer et al., 2006). Moreover, distance testing tends to run into unexpected problems that complicate the process (Kaikkonen et al., 2005). The second mentioned method appears as preferable for many cases. For laboratory testing, users are located centrally in one place. Hardware and software testing requirements are thus secured by equipment in the test lab.

The complexity of the gathered data is constantly rising as new testing techniques are being implemented. A visual form, such as audiovisual records, is a typical style of output when performing a usability testing in laboratory. Analysis of such data is highly time-consuming (Prasse, 1990). In contrast of quantitative data, it is not yet possible to algorithmically analyze the audiovisual outputs or transform them into such form. Therefore, researchers have to perform a thorough review and analysis of the recordings. When testing higher amount of users and tasks, the time consumption of the analysis is astronomical (Nørgaard et al., 2006).

When testing usability of applications, it is often needed to analyze users' behavior in terms of identifying their activities. The activity may be that the user is working on the assignment without problems, is searching for something, is absolutely lost in UI, is filling form, is studying the manual, etc. The outputs then can be easily visualized using various technics such as pie chart. The pie chart has a significant predictive value and can be easily understood by management workers. The time consumption of the user activities is important, for example, in decision making about the usage of the tested software (Harel et al., 2008). The Department of Information Technologies at CULS in Prague used this type of output to support various usability research studies for the Czech Ministry of Agriculture and the Ministry of the Environment. This type of analysis usually needs to employ a log of the desired user activities. This can be achieved by the pre-analysis of the user activities during the test followed by detailed tagging of the testing video recording (Wynn and Still, 2011). However, this type of analysis has to be done manually by the usability experts. It is a very time-consuming activity. Some usability test may last dozens of minutes. Many parts of the video recording need to be reviewed repeatedly. One of the possible solutions to this critical problem is some kind of automatic processing of the video recordings, and auto-tagging of the user activities (Petropavlovskiy and Nefedova, 2017).

Recently, there have been a lot of new advanced methods and technologies for usability testing in a laboratory. The gathered data are highly valuable for understanding users' behavior (Çakar et al., 2017). Besides the audiovisual recordings, it is possible to track the user input (e.g. click tracking and keyboard tracking). It is possible to visualize and analyze the user behavior using eye tracking equipment (Hild et al., 2012) which can provide additional valuable insights (Wang et al., 2019). Moreover, it is possible to get various biometrical data. These data can enrich the overall picture of users' experience such as stress or emotions (Mirza-Babaei et al., 2011). In regard to such data volume and its complexity, there is a problem with the processing and transformation into a necessary form for testing hypotheses and drawing conclusions. Manual processing of such data is very difficult and often not suitable for real scenarios.

The objective of the paper is to introduce a methodology for the research to find methods and algorithms for automatic data processing. The output should be a correct identification of the user's activities during usability testing. The research is currently conducted at the Department of Information Technologies at CULS Prague.

2. Methodology

Currently, self-developed software is used to analyze the video record and tag the user activities, named VEA-UX. It supports definition of custom activities, option to save the project for later work, and dividing tagger's workload into sections. The sections mostly serve as checkpoints for the tasks which users have to perform during the testing. The application works on a slightly different principle than other similar software. The events are marked when they conclude. This approach helps to save time needed for the analysis. It avoids excessive rewinds to the beginning of the activity. The software also supports backward editing of any tagged events and sections. The capabilities of the software are illustrated in a Figure 1. Despite the fact, that the VEA-UX application can save time for the analysis, it is still time-consuming. Nevertheless, the VEA-UX will be used later for evaluation.

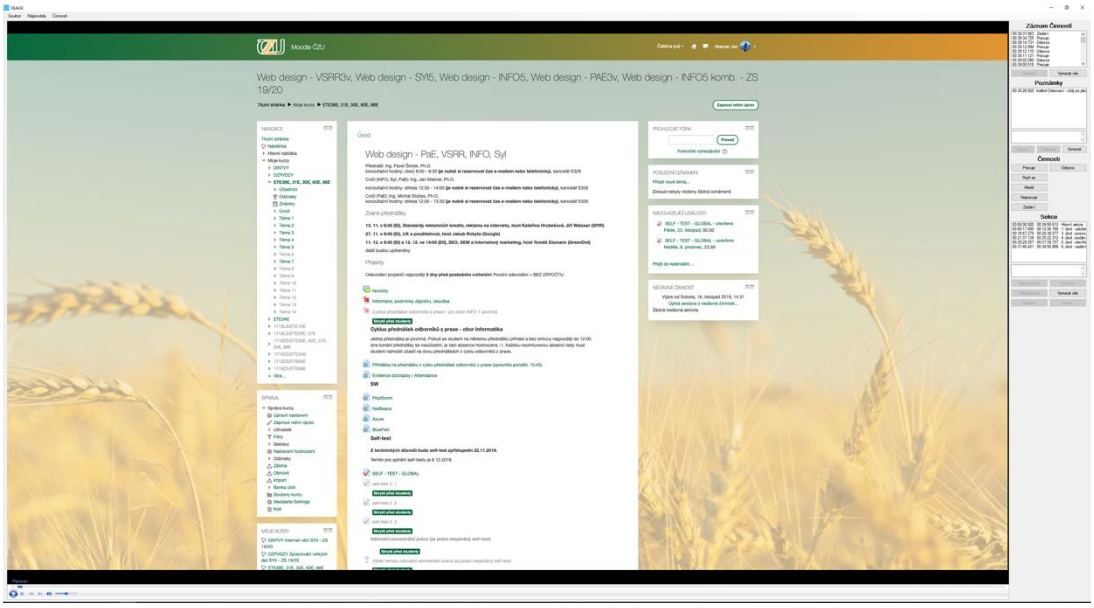

Figure 1. VEA-UX software used for tagging user activities from a video recording

Experiments

The proposed methodology consists of several steps. First of all, some experiments were carried out. At the Faculty of Economics and Management, specialized laboratory for usability testing (HUBRU – Human Behavior Research Unit) was used. The laboratory is equipped with computers, devices for eye tracking, and advanced recording equipment.

For the first set of experiments, testing environment in Moodle was created. It consisted of a copy of the existing course taught by the Department of Information Technologies. The testing scenario consisted of 5 tasks, ranging from very easy to difficult to accomplish. The tasks for users were as follows:

1. Find and write down at what time the Lectures are held.
2. Find out who attends the course with you. Choose one of your fellow students and send him a message.
3. Hide or show some item from a course.
4. Create a new topic in the course. Add to the topic some file material.
5. Change overall score (grade) of a student named Michael White to 200 points.

Eight users were tested to get the first results. The test group contained users with no experiences with Moodle as well as students with some knowledge about its functions, and experienced teachers. To gather the data for later analysis, the Tobii Pro X2 (60Hz version) eye tracker together with click and keyboard tracking, and Huawei Watch 2 were used. A very important setup which needs to be done before such testing, is to synchronize the times across devices. The synchronization is necessary to properly merge the data from different sources. We also placed a smartwatch on users' dominant hand (which they use to control the mouse). This setup optimizes the use of accelerometer and gyroscope data.

Data preparation

The eye tracking data together with click and keyboard tracking data were gathered through a Tobii Pro Studio software which operates the eye tracker device. The timestamp was retrieved from the computer time. Data from smartwatch are retrieved via the application SensorCap. It allows to synchronize the time with a custom NTP (Network Time Protocol) server and easily export sensor data in CSV format. The same NTP server has to be set on the computer with eye tracking. Selected sample data are shown on Table 1 – eye tracking, and in Table 2 – smartwatch.

Table 1. Sample data from eye tracking (selected columns)

Recording Timestamp	Local Timestamp	GazeEvent Type	GazeEvent Duration	GazePoint. Index	…
23	15:18:42.210	Fixation	400	1	…
39	15:18:42.226	Fixation	400	2	…
57	15:18:42.244	Fixation	400	3	…
73	15:18:42.260	Fixation	400	4	…
90	15:18:42.277	Fixation	400	5	…

Table 2. Sample data from smartwatch - heartrate

eventTimeNanos	deviceTimeMillis	ntpTimeMillis	sensorAccuracy	HTR
7464907545779	1548771430329	1548771430044	3	91.0
7465931257028	1548771431353	1548771431068	3	91.0
7466929089111	1548771432351	1548771432066	0	92.0
7467999088174	1548771433421	1548771433136	0	93.0
7468908427913	1548771434330	1548771434045	0	93.0

The data shown in Table 1 and Table 2 need to be joined together. Additionally, there are more smartwatch data sets. Firstly, we used data from accelerometer sensor. The data have similar structure as shown in Table 2. To sum up, we needed to merge three data sets – eye tracking with keyboard and click streams, heart rate data, and accelerometer data. The problem is that the granularity is different – the values are taken at different periods. Using Jupyter notebook and Python programming language we developed an algorithm to merge the datasets. Using the smallest granularity according to Local Timestamp in Table 1, the data were resampled into milliseconds. Subsequently, all three datasets can be merged and grouped.

3. Further Research

Data Analysis

Firstly, the exploratory analysis will be carried out. Correlation analysis will be used to determine whether there are variables that can be omitted. Visualization also helps in understanding the data. We use mostly Tableau. For the later model development, we plan to use machine learning. Therefore, we will firstly engage statistical classification using artificial neural networks. The TIBCO Statistica software can be easily used without further programming.

During the analysis, the input data sets will need to be adjusted. Several factors need to be considered. The input variables are the most significant for the result. Other variables can be computed, for example delayed values can help to recognize longer activities.

The data, which has been gathered, can serve as a good basis for further research. Very important factor is to properly prepare the testing environment as well as the tasks for the users. Additional tests with users will be done according to the data analysis. We have other smartwatches available. It is also possible to use dedicated equipment for biometric feedback. It can measure additional values such as skin temperature, skin conductance, motility, and advanced pulse data.

4. Conclusions

When finished, the proposed research and its output methods can help to save time during usability research studies. The methods and algorithms will be able to automatically identify the problem parts of the tested software. This can help, for example, in backwards analysis of the video records to identify the problem parts. Usability experts can then focus only on the most important aspects of the tested product. In some cases, it is required to analyze the effectivity of using the product. We can identify, when users spend too much time by searching, navigating, or by other activities which are not productive. The results of such usability studies can be delivered in a radically shorter time.

The proposed research can be generalized to usability of other then software products. For the tracking of eye movements, it is possible to use special glasses such as *Tobii Pro Glasses*. The user does not have to be at a fixed position in front of the computer or other device. For example, farmers can wear the glasses when operating a tractor, harvester, etc. Together with the use of smartwatches, users can wear the equipment for a longer period of time and in a real, natural environment. The testing would not be affected by any interference or stress caused by the unnatural environment in a laboratory. To analyze testing of such length, the use of traditional video analysis methods would be time-consuming, economically inefficient, and therefore ill-advised.

Acknowledgments

The results and knowledge included herein have been obtained owing to support from the following institutional grants. Internal grant agency of the Faculty of Economics and Management, Czech University of Life Sciences in Prague, grant no. 2019B0009 – “Life Sciences 4.0“.

References

Atterer, R., Wnuk, M., Schmidt, A., 2006. Knowing the user’s every move. https://doi.org/10.1145/1135777.1135811

Benda, P., Šimek, P., Masner, J., Vanek, J., 2017. Analysis of eAGRI web portal ergonomics and presentation of information in terms of the general public. Agris On-line Pap. Econ. Informatics 9. https://doi.org/10.7160/aol.2017.090401

Çakar, T., Rızvanoğlu, K., Öztürk, Ö., Çelik, D.Z., Gürvardar, İ., 2017. The use of neurometric and biometric research methods in understanding the user experience during product search of first-time buyers in E-commerce, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-319-58634-2_26

Harel, A., Kenett, R.S., Ruggeri, F., 2008. Decision support for user interface design: Usability diagnosis by time analysis of the user activity, in: Proceedings - International Computer Software and Applications Conference. https://doi.org/10.1109/COMPSAC.2008.61

Hild, J., Peinsipp-Byma, E., Klaus, E., 2012. Improving usability for video analysis using gaze-based interaction, in: Full Motion Video (FMV) Workflows and Technologies for Intelligence, Surveillance, and Reconnaissance (ISR) and Situational Awareness. https://doi.org/10.1117/12.919042

IDF, 2019. Usability [WWW Document]. URL https://www.interaction-design.org/literature/topics/usability

Jarolimek, J., Masner, J., Vanek, J., Pankova, L., 2019. Assessing Benefits of Precision Farming

Technologies in Sugar Beet Production. List. Cukrov. A Repar. 135, 57–63.

Kaikkonen, A., Kallio, T., Kankainen, A., Cankar, M., Kekalainen, A., 2005. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. J. USABILITY Stud.

Mirza-Babaei, P., Long, S., Foley, E., McAllister, G., 2011. Understanding the contribution of biometrics to games user research. Proc. DiGRA.

Nørgaard, M., K, H., K, 2006. What do usability evaluators do in practice? An explorative study of think-aloud testing. In: Proceedings of the. DIS 2006 Conf. Des. Interact. Syst. ACM Press. New York. https://doi.org/10.1145/1142405.1142439

Petropavlovskiy, M., Nefedova, O., 2017. AUTOMATIZATION OF COLLECTION AND USE OF THE INFORMATION FROM OFFICIAL WEBSITES OF EDUCATIONAL ORGANIZATIONS. IJAEDU- Int. E-Journal Adv. Educ. https://doi.org/10.18768/ijaedu.336265

Prasse, M.J., 1990. Video analysis method. An integrated approach to usability assessment, in: Proceedings of the Human Factors Society. https://doi.org/10.1177/154193129003400436

Rubin, J., Chisnell, D., 2008. Handbook of Usability Testing, Indianapolis, IN: Wiley Pub. https://doi.org/10.1007/s13398-014-0173-7.2

Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., Fieldman, A., 2019. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. Int. J. Hum. Comput. Interact. https://doi.org/10.1080/10447318.2018.1464776

Wynn, J., Still, J.D., 2011. Motivating change and reducing cost with the discount video data analysis technique, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-21708-1_37