Challenges in Usability Evaluation of Expert Domain Products

Johanna Viitanen, Sami Karjalainen, Heli Kautonen, Anna Laukkanen

Software Business and Engineering Institute
Helsinki University of Technology
P.O. Box 9210, 02015 TKK, Finland
johanna.viitanen@hut.fi, sami.karjalainen@korento.net,
hkautone@cc.hut.fi, aklaukka@cc.hut.fi

Abstract

Evaluating the usability of products that are meant for professionals of a specific domain is a demanding task. Evaluators coming from another field of expertise need to confront unfamiliar use cases and user needs. This article describes the challenges by introducing two usability evaluation cases. The first product under evaluation was an optimization software for mass haulage projects used in highway and railroad construction, and the second a touch-screen prototype of an anesthesia monitor for healthcare professionals. Evaluation processes were conducted using best-practice methods, viz. heuristic evaluation and user testing. The results indicate that the biggest challenges are related to defining usability criteria, choosing appropriate methods and conducting user tests. Usability evaluation will provide better results, if evaluators can study the purpose of the product in actual use situations already at the beginning of the evaluation process.

1 Introduction

Expert domain products are used by people who are professionals in some specific domain area. These products are used to aid work performance. The emphasis is on the task to be accomplished, not on the tool, which processes and presents task related data. Complexity and a wide range of features are characteristics of an expert domain product. Evaluating usability of such complex products is challenging. Expertise in usability does not carry along the ability to comprehend the domain nor the work tasks. Evaluators face challenges when defining the usability criteria and choosing evaluation methods, although there are well established practices for both.

1.1 Usability evaluation

Expert review is an easy and efficient way to provide usability data. An experienced researcher can easily conduct a heuristic evaluation [1] that brings impressive results. However, domain specific usability problems cannot be revealed if the evaluator is not familiar with the work itself and the user's cognitive environment [5]. Another expert review method, viz. cognitive walkthrough, fails in capturing the overview of complex applications since it focuses on single tasks [3].

Usability evaluation obviously requires end-user participation. User tests have shown their ability to track user needs for a product, and especially for the user interface [1]. Still, user tests seldom uncover tacit usability problems, which are often the most significant ones. The thinking-aloud method is often used to reveal the user's cognitive activity. Others consider it unnatural for a domain professional to verbalize task decisions and own performance [5]. Furthermore, the test situation usually gives a fragmented view of the work, which does not correspond to the real life.

In addition to the criticism against usability evaluation methods, usability studies as such have been considered inapplicable for practice [4]. Usually the purpose of usability studies is to trace usability problems, whereas the design team would mainly need fresh ideas for product development. Thereby, usability evaluators must consider various aspects of the target phenomenon, the product, as well as the context of use.
1.2 Aim of paper

This paper describes challenges in the usability evaluation of two expert domain products, and inspects how research methods and their conduction affected the results. Furthermore, the aim is to detect shortcomings of the methods used and, thus, reconsider the results. Biggest challenges were related to finding the most appropriate methods for evaluating a product from an unfamiliar domain, and engaging domain experts to user tests. In order to prepare user tests and find appropriate test users, evaluators were highly dependent on the client who was also the product developer. The results are therefore influenced by the co-operation with domain expert users as well as the client.

The primary research question was:
- What are the challenges of evaluating the usability of an expert domain product?

The question was then approached with further questions:
- How should the unfamiliar domain be considered?
- How should evaluators approach experts of a domain as product users?
- What should be the role of the client?

The paper describes the challenges by presenting two cases where some best-practice usability methods were used. First there is and introduction to the products and evaluation methods. Then these two are weighed up together. The paper concludes with a discussion of challenges and a list of guidelines for similar cases.

2 Evaluated products and usability evaluation methods

Usability of two expert domain products was evaluated. The products were an optimization software for mass haulage projects used in highway and railroad construction, and a touch-screen prototype of an anesthesia monitor for healthcare professionals. Although the products differ considerably from each other, they can both be classified as expert domain products, and the chosen usability evaluation methods were quite similar. In this section, the products and the usability evaluation methods are described.

2.1 Optimization Software for Road Construction

The optimization software can be used in all road construction project steps, from pre-tendering planning to project implementation and control. It is used in demanding highway and rail construction projects.

Typical users are road construction professionals, who work in planning and supervision. They have a technical background and use computers daily, but are typically not expert users of computers. They may use optimization software intensively during certain phases of a construction project, but not continuously, and not in every project.

The goals of the usability evaluation were to study the usability of a new version of the software and to give proposals for improvement. The main interest was to find, if it is possible to use the program without extensive user training. An important objective was also to look for problems in the software that could lead to critical errors.

2.2 Anesthesia Monitor

The anesthesia monitor is used in hospitals to control the patient’s state of health during operation. The most typical function of the monitor is to make different types of measurements of the state of health, measuring for example heart rate (ECG), oxygen saturation, and blood pressure. The monitor also functions as an alarm device and gives audible and visible warnings of possibly dangerous changes in the patient’s state of health. The monitor collects and saves numerical data for later viewing and printing.

The users are professionals in their field of expertise. Their experience as users of an anesthesia monitor varies. Different types of monitors exist, so the users may have practiced the use of an anesthesia monitor with any of the several types of monitors.
The evaluated prototype was a touch-screen version of the monitor. Users navigate in menus by touching menu items, and set values by touching sliders.

The main objective of the usability evaluation was to analyze how the usability of the anesthesia monitor evolves with the new touch-screen technology. The evaluation and the suggestions for development were focused on improving efficiency and satisfaction of use and reducing the number of errors.

2.3 Usability Evaluation Methods

Both products were evaluated by similar methods. Usability evaluation started with heuristic evaluation, which was based on ten usability heuristics [2]. Next, usability tests were conducted in a usability laboratory. The evaluation of the anesthesia monitor was complemented with a heuristic evaluation by an expert of both the domain area and usability.

Both products were tested with four test users. Usability tests were based on test tasks, which were given to the users one at a time. The tasks were planned to be as close as possible to real work tasks. The test users were experts of the domain area, actual or potential users of the products. They were instructed to talk aloud during the test. Usability tests were videotaped. Material was also collected by interviewing the test users about their work and the test. Additionally, the test users filled short questionnaires.

3 Challenges of evaluation

While evaluating expert domain products, usability specialists need to give their opinion on a product not familiar to them. For each phase of the evaluation to give valuable information, the methods and the progression of the evaluation should be carefully planned. The results of and the related proposals for improvement clearly depend on how well the evaluators have understood the purpose of the product and the work process it should be supporting.

3.1 Planning the evaluation

If the domain is unfamiliar and complicated, the evaluators should reserve more time than usually for the evaluation. Getting to know the product and its use increases the workload. Support from the client is needed while preparing the evaluation: understanding the product, finding users for the tests and making test tasks cannot be accomplished without help. Cooperation is especially indispensable at the first stages of the process.

Since time resources are always limited, it is typically not possible for usability evaluators to get to know perfectly an expert domain product. Thus, it has to be defined, which parts of the product are included in the evaluation. The definition should be based on the goals of the evaluation, taking also into account the resources of the evaluators. Besides, the stage of development of the product affects the ways in which evaluation can be conducted. Finally, it is important that the evaluators and the client have come to an agreement over the limits of the evaluation.

3.2 Expert review

Although doing an expert review, such as heuristic evaluation, is challenging, when the domain is unfamiliar, it supports the evaluation process in several ways. The timing of the expert review in relation to the whole process can affect the results found using it. If it is performed at an early stage, it helps the evaluators to get to know the product. In addition, problems can be detected, which would be left unnoticed by people already familiar with the system. On the other hand, it is hard to get reliable and comprehensive results when the evaluators do not have a broad view of the product.

The results of the expert review provide material for preparing the usability tests: areas considered problematic in the expert review can be covered in the test, and comparing the results helps to define the seriousness of the problems found. Still, the most important results regarding the whole evaluation process come from the usability tests.
### 3.3 Usability tests

Usability tests are carried out with people belonging to the actual user group of the product. To get in touch with users, help from the client is often required. The users affect also the choice of evaluation methods: if they are difficult to reach and have busy schedules, methods calling for several users, such as the focus group method, cannot be chosen.

Since the users are found through the client, the name and the products of the client can be familiar to them. It is also possible that they have earlier used similar products of different manufacturers. The potential preconceptions caused by previous knowledge should be taken into account while analyzing the results.

Even though the evaluators have to be acquainted with the product before the tests, they probably will not get profound enough knowledge of the users' tasks to prepare test tasks. Thus, the client needs to participate also in making test tasks and model solutions and instructions for them.

When the system is unfamiliar, the instructor of the test faces several challenges. The users may ask questions, and many types of problems can arise when going through the tasks. The instructor should therefore study the tasks before the test and find potentially difficult parts and dead ends. It is a good idea to create files that bring the system to various stages of the test tasks, so that the test can be continued from any chosen point. The importance of a pilot test prior to the actual tests is elevated, since it allows the evaluators to test and practice the test situation.

Conversation between the instructor and the user is an essential source of information when testing expert domain products. The test tasks act as a basis for discussion between the expert user and the instructor – this is their most important function in the test, above that of directly detecting usability problems. The instructor should address issues that emerge during the test immediately and ask more questions on them, even if this slows down the test. Conversation taking place outside the actual test situation, for example during a coffee break, can also bring new information.

The fact that the instructor has only little knowledge on the domain helps creating a natural test situation: the relation of the instructor and the expert user resembles that of a master and his apprentice. Consequently, the method of thinking aloud suits well the test situation, as the user is explaining the problems caused by the system and the related work tasks to a person less familiar with the subject.

### 3.4 Goals of evaluation

The goals set at the beginning of the evaluation cannot always be met completely. Not all usability criteria can even be measured at each phase of development of the product. It is important to choose criteria that can be tested and are relevant to the development of the product.

Evaluating efficiency is difficult, if not impossible, if the prototype is incomplete. In the evaluation of the anesthesia monitor, one of the goals was to measure efficiency against the current product. This was only partly successful, since the prototype was still under development. A reliable evaluation of efficiency is also impossible, if the users see the product for the first time.

The evaluation of user satisfaction is also difficult when using an unfinished prototype of the product. Besides, the users' experiences of the client and of similar products affect their attitudes. For these reasons, the users' opinions on the anesthesia monitor differed considerably. In the evaluation of the optimization software, the users had a positive view of the client, which may have reduced their willingness to criticize the product.

A characteristic of expert domain products is that errors caused by the user interface can have serious consequences – at worst lives may be lost (anesthesia monitor) or the company may suffer large economic losses (optimization software). The probability of serious errors is hard to estimate, when the domain is unfamiliar and complex. Work in an expert domain involves facing completely new products only infrequently. Thus, the importance of evaluating learnability in expert domain products is in many cases questionable. It would possibly be more useful to focus to usability characteristics essential in everyday use, such as efficiency and satisfaction.
On the other hand, there are reasons, which make a certain degree of intuitiveness and learnability significant also to an expert domain product. According to the developer of the optimization software, training is rarely bought, and users typically familiarize themselves with the system alone or with the help of colleagues. The users of the anesthesia monitor are trained to use a monitor, but, when in a work situation, they may have to use a monitor not exactly similar to the one with which they have practiced. Thus, intuitiveness and accordance to standards and conventions of the domain are important.

Besides testing the usability criteria considered important for each product, an important objective of the usability evaluations was to find ways to improve the products and solve problems encountered during the evaluation. When developing suggestions for improvement, it is of great benefit if a domain expert can be consulted. This can be done for example by organizing a review, in which the suggestions are evaluated.

4 Conclusion

Two usability evaluation cases showed the challenges of evaluating an expert domain product. The evaluators faced many challenges, because the target domain was unfamiliar to them. Firstly, it was problematic to define viable usability criteria and metrics, since evaluators had no insight into the context of professional use. For example, the review and test situations with the anesthesia monitor seemed to underline the importance of intuitiveness, whereas some other criteria, such as effectiveness or a low rate of errors, may be more critical in live use situations. Tracking real usability problems in short-term usability tests was also considered difficult. For example, how can one measure errors that accumulate in weeks or months of use? The results certainly excluded some underlying usability problems.

Still, examining the usability of such products is always valuable. Case projects were conducted using best-practice usability evaluation methods: heuristic evaluation and user testing. In spite of obvious challenges and shortcomings, they provided useful information of apparent usability problems. Thus, they can be recommended for evaluating expert domain products. However, better results may be obtained if evaluators understand the purpose of the product and have the possibility to study users in actual work situations. This should be done already at the earliest phases of the evaluation process.

Guidelines for the usability evaluation of expert domain products:

- When planning time and resources, foresee the time required for learning the use of the product.
- Focus the evaluation on limited features of the product.
- Study the context of use by spending time on the site, i.e. join, observe, and interview users in their real work environment.
- Be prepared for dead-ends during the test.
- Find ways to generate conversation with users about their real work and tasks.
- Take into account conventions and tacit knowledge of the target domain.

5 Acknowledgements

Referred usability evaluation studies were conducted on Usability Evaluation course at the Helsinki University of Technology. We thank Sirpa Riihiaho and Laura Turkki from the Software Business and Engineering Institute, SoberIT at the Helsinki University of Technology, Clinical Systems of GE Healthcare IT in Helsinki and Dynamic System Solutions Inc.

References

