Usability Evaluation of Interactive Systems using TREVIS

Michael Hähnel, Nico Hamacher, and Karl-Friedrich Kraiss
Department of Technical Computer Science, Aachen University of Technology, Ahornstr. 55, D-52074 Aachen, Germany, E-mail: {haehnel, hamacher, kraiss}@techinfo.rwth-aachen.de

Abstract

While the usability of interactive systems is getting increasing attention, the designing engineer has to evaluate the specifications and implementations of the developed system. To support these evaluation throughout all phases of the whole development process the tool TREVIS is being implemented and will be described in this paper. Based on the GOMS-theory TREVIS uses formal, normative models to predict measures to estimate the usability of the examined system. It supports the engineer with building these models either manually or automatically and analyses these models as well as empirical obtained action logs.

1 Introduction

Today, usability engineering of interactive systems is getting increasing attention. In general usability is determined by testing prototypes. The disadvantage of this approach is that these tests can be undertook only in late stages of development process. An early analysis of usability would be a significant advantage with regard to saving time and resources (Kraiss, 1995). Furthermore an early analysis has to generate meaningful measures, like predictions of the execution and learning time, the complexity and consistency of the given device. Such robust measures can be obtained by using the tool TREVIS (**T**ool for **R**apid **Ev**aluation of **I**nteractive **S**ystems), which is presented in this paper. TREVIS enables the design engineer to model the behaviour of an user while interacting with a device and derive usability measures from this simulation (Marrenbach, 2001).

2 Usability Evaluation

One method of usability evaluation is a formal evaluation method which can be undertaken in early phases of system development. This formal evaluation method is developed by Card, Moran and Newell (1983) and describes the interactions between a user and an interactive system. It is called GOMS, which is an abbreviation for the components of the model: Goals, Operators, Methods and Selection rules. Goals describe the given tasks, whereas Methods represent the corresponding subtasks. Operators model the single sensorical, cognitive and motorical actions, the user has to perform to reach the goal. Selection rules enable the engineer to model different ways to reach a goal or subgoal. A GOMS model is also named user model. An analysis based on GOMS generates qualitative as well as quantitative predictions, like execution and learning time (Kieras, 1988). The execution time describes the time to reach the goal whereas the learning time specifies how much time an operator needs to learn the whole task. In addition, the consistency of the device, which describes the uniformity of operation, can be analysed. Although the use of GOMS models is not very complicated (John & Kieras, 1997), it is very tedious to build these models manually. Hence, a tool is needed, which integrates the GOMS theory in the development process and supports efficiently creation and analysis of user models.

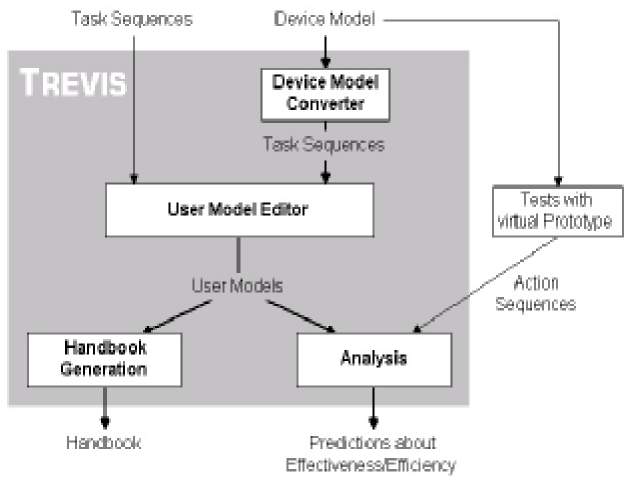

Figure 1: Architecture of TREVIS

3 System Architecture

The tool TREVIS (**T**ool for **R**apid **Ev**aluation of **I**nteractive **S**ystems) supports the evaluation of interactive systems in various ways. TREVIS includes four main modules, depicted in figure 1 (Hamacher & Marrenbach, 2001):

- Based on task sequences as one result of the requirements the user models can be created manually in the user model editor. The tool supports this process e.g. by offering a library for reusing components and a graphical editor. In TREVIS a whole project management is also implemented, where the project represents the interactive system and the user models describe the tasks which have to be done.
- The device model contains details about the inner works of the device. If a device model was created in an earlier stage of the development process, the task sequences can be generated semi-automatically using the device model converter. The user models can be created also from these sequences. Import filters for SDL are implemented (Hamacher *et al,* 2002).
- In the analysis module different analysis are included, which depend on the development phase, where TREVIS will be used. These analysis are described in the next section.
- The user models contain the complete description of the procedural knowledge that the user has to know in order to perform tasks using the device. Hence, a handbook based on the user models can be created by the handbook generator (Elkerton & Palmiter, 1991).

4 Results of the different analyses

The analysis module of TREVIS generates various outcomes, which depend on the kind of used analysis and the development phase, TREVIS will be used in. Five different types of analysis are realised and will be described in the following subchapters.

- **User Model Analysis**
  The user model analysis generates qualitative as well as quantitative predictions, like execution and learning time (as already described with NGOMSL). This analysis can be made in early stages of the development process without a real existing system or prototype.

- **Design Analysis**
  A comparison between different user models is implemented in the design analysis module, which can be used as a basis for design decisions. Beside a suitable representation of the user model analysis of all included user models, the used operators are diagrammed for each user model to compare the actions, which come into operation.

- **Consistency Analysis**
  Based on several user models, the consistency analysis compares the reuse of menus and actions and generates measures about the similarity of operation (Hamacher & Hähnel, 2001). Operator sequences in the user model (i.e. methods) are compared and similarity coefficients are calculated by the ratio of the number of different operators to the number of operators in total. Re-usage of similar operator sequences lead to higher similarity in using the system and will lead to better similarity coefficients. Additionally the coefficients are used to calculate a consistency measure that estimates the degree of consistency of the entire system.
  This analysis can be applied to the full functionality of a system as well as to parts of a system in which consistency of functionality is an important demand.
  This analysis also allows to compare either different designs or several systems of one kind in terms of consistency.

- **Action-sequence Analysis**
  In the action-sequence analysis, action sequences resulting from testing a prototype can be imported and analysed. A grouping of different sequences is possible, e.g. to perform an analysis of significance or an analysis of variance. With this feature, TREVIS is also applicable in late stages of a development process. Away from normative models this analysis allows an estimation of the usability by considering a prototype and subjects.

- **Goal Analysis**
  The goal analysis compares these action sequences with the user models. This analysis shows the differences between the actions specified in the user models and the activities, the users performed while interacting with a prototype. Various figures were calculated, e.g. action frequency, error rates, number of aborted action logs. These outcomes are displayed in various graphical and alphanumerical ways. This analysis gives hints where a system has to be re-designed to improve the intuitional usage.

5 Applications

To show the usefulness and the reliability of TREVIS an evaluation of three mobile phones was undertaken. Therefore 12 representative tasks (i.e. “calling a number”, “calling a person by using the phone book”) were chosen, the corresponding user models were created and analysed.
Table 1 shows an example of the predicted execution times calculated by TREVIS as well as execution times determined by a set of action logs which were generated by empirical experiments (Ziefle, 2002). These are the results of the representative task “Switch on the transmission of the phone number”.
The predicted execution times and the average

execution times of the action logs vary between 11,38% and 67,07% depending on the distribution of times of the action logs, which varies intense because action logs of novices and experts are included. That is why the error frequency is quite high.

Table 1: Goal analysis of a representative task on three mobile phones

mobile phone	TREVIS predicted exec. time	action logs (average)	empirical standard deviation (action logs)	variance (in %)	error freq. (in %)
Nokia 3210	17,81 s	19,84 s	8,63 s	11,38	53,33
Motorola 7350	13,76 s	22,99 s	10,82 s	67,07	115,24
Siemens C35	19,35 s	27,07 s	14,31 s	39,90	73,68

Table 2 shows the calculated consistency measures applied on a selected set of representative tasks. The better value of consistency of the Nokia mobile phone in comparison to the Siemens and Motorola was evaluated by empirical experiments and reflects the results of a questionnaire applied to these mobile phones.

Table 2: Consistency measure of three mobile phones

mobile phone	consistency
Nokia	0,4177
Siemens	0,4010
Motorola	0,3910

6 Summary

This paper presented the tool TREVIS which offers the possibility to evaluate an interactive system throughout its development process.

In early stages of the development process user models, which describe the interactions between an operating user and an interactive system, are used to analyse a system's efficiency and consistency. The designing engineer is supported in building this models either manually or semi-automatically from specification data. By importing action-sequences from interactions with a prototype, a system can also be evaluated in later phases of the development process. Several analyses based on both kinds of input data can be performed, i.e. the comparison between action-sequences and related user models.

The tool was evaluated with different kinds of interactive systems; results from an evaluation with mobile phones were presented. TREVIS proved its usefulness with supporting the developing engineer in evaluating and re-designing the system throughout the whole development process.

7 References

Card, S., Moran, T., and Newell, A. (1983): *The psychology of human computer interaction,* Lawrence Erlbaum.

Elkerton, J, and Palmiter, S (1991): *Designing help using a goms model: An information retrieval evaluation*, in "Human Factors", Volume 33, pp. 185-204.

Hamacher, N., Kraiss, K.-F., and Marrenbach, J. (2002): *Einsatz formaler Methoden zur Evaluierung der Gebrauchsfähigkeit interaktiver Geräte*, in "IT+TI Informationstechnik und Technische Informatik", Volume 44, Heft 1, pp. 49-55, Oldenbourg-Verlag.

Hamacher N., and Marrenbach J. (2001): *Analytical Evaluation of Interactive Systems regarding the Ease of Use*, in Stephanidis, C. (Eds.): *Proceedings of HCI International 2001. Universal Access in HCI. Towards an Information Society for All*, Volume 3, pp. 585-589, August 5-10, New Orleans, Lawrence Erlbaum Associates.

Hamacher N., and Hähnel M. (2001): *Konzept für die automatische Generierung von Komplexitätsmaßen zur Evaluierung interaktiver Geräte*, in Gärtner, K.-P. and Grandt, M. (Eds.): *43. Fachausschusssitzung Anthropotechnik der DGLR*, Volume DGLR-Bericht 2001-06, pp. 117-127, 23.-24. Oktober 2001, Hamburg, DGLR Bonn.

John, B.E., and Kieras, D.E. (1997): *Using GOMS for user interface design and evaluation: Which technique?*, in "ACM Transactions on Computer-Human-Interaction", Volume 3, pp. 287-319.

Kieras, D. (1988): *Towards a practical GOMS model technology for user interface design*, in Helander, M. (Ed.): *Handbook of human-computer interaction*, pp. 135-157, Amsterdam: North Holland.

Kraiss, K.-F. (1995): *Modellierung von Mensch-Maschine Systemen*, in Willumeit H.-P. (Ed.): *Verläßlichkeit von Mensch- Maschine Systemen*, ZMMS - Spektrum, Volume 1, pp. 15-35, Berlin.

Marrenbach, J. (2001): *Werkzeug-basierte Evaluierung der Benutzungsfreundlichkeit interaktiver Endgeräte mit normativen Benutzermodellen*, Dissertation, Technische Informatik, Shaker-Verlag.

Ziefle, M. (2002): *Usability of menu structures and navigation keys in mobile phones: A comparison of the ease of use in three different brands.* Paper presented to the 6th International Conference WWDU-Work with Display Units 2002 (this conference).