Usability Evaluation of Tools for Nomadic Application Development

Cristina Chesta (1), Carmen Santoro (2), Fabio Paternò (2)

(1) Motorola Electronics S.p.a. – GSG Italy
Via Cardinal Massaia 83, 10147 Torino Italy
Cristina.Chesta@motorola.com

(2) I.S.T.I. - C.N.R.
Via G. Moruzzi 1, 56100 Pisa Italy
{c.santoro, f.paterno}@cnuce.cnr.it

Abstract.

Evaluating the usability of tools for development of nomadic applications introduces specific issues since it requires attention to both the interface of the tool itself, and the interface of the application produced through the tool. The paper addresses this problem by discussing the criteria and methodologies applied as well as the results obtained in an experimental activity on the subject.

Introduction

Several research activities have addressed design principles, methods and software development tools to increase systems usability while reducing development costs. Meanwhile a number of criteria and methodologies to evaluate the usability have been proposed. However, there is a lack of studies directed to evaluate the usability of the development tools, in particular when nomadic applications, which can be accessed through different types of interaction platforms, are considered. This requires a double point of view, on the tool itself and on the product realized through the tool.

This problem is acquiring an increasing importance because many organizations have to develop new applications able to exploit the possibilities provided by mobile devices. This paper addresses such issues, by presenting some results of the evaluation activities carried on within the framework of the CAMELEON project in Motorola GSG Italy with the objective to assess the usability of the TERESA tool for design and development of multi-platform applications proposed by the HCI Group of I.S.T.I.-C.N.R. In particular, in the paper at first the TERESA tool is introduced and described, then the evaluation criteria adopted and the experiment organization are presented, finally some significant results followed by the concluding remarks are provided.

The TERESA tool

TERESA is a transformation-based environment designed and developed at the HCI Group of I.S.T.I.-C.N.R. It is intended to provide a complete semi-automatic environment supporting a number of transformations useful for designers to build and analyse their design at different abstraction levels and consequently generate the concrete user interface for a specific type of platform.

The abstraction levels considered are: the task model level, where the logical activities to support are identified; the abstract user interface, in this case the interaction objects (but classified in terms of their semantics, still independent from the actual implementation) are considered, and the concrete user interface (the actual corresponding code). The main transformations supported in TERESA are:

- *Presentation sets and transitions generation.* From the specification of a ConcurTaskTrees [4] task model it is possible to obtain the set of tasks, which are enabled over the same period of time according to the constraints indicated in the model. Such sets, depending on the designer's application of a number of heuristics supported by the tool, might be grouped together into a number of *Presentation Task Sets (PTSs)* and related *Transitions* among the various PTSs.
- *From task model -related information to abstract user interface.* Both the task model specification and PTSs are the input for the transformation generating the associated abstract user interface, which will be described in terms of both its static structure (the "presentation" part, which is the set of interaction techniques perceivable by the user at a given time) and dynamic behaviour (the "dialogue" part, which indicates what interactions trigger a change of presentation and what the next presentation is). The structure of the presentation is defined in terms of elementary interactors characterised in terms of the task they support, and their composition operators. Such operators are classified in terms of the communication goals to achieve: a) *Grouping*: indicates a set of interface elements logically connected to each other; b) *Relation*: highlights a one-to-many relation among some elements, one element has some effects on a set of elements; c) *Ordering*: some kind of ordering among a set of elements can be highlighted; d) *Hierarchy*: different levels of importance can be defined among a set of elements.
- *Automatic abstract UI generation.* Through this option the tool automatically generates the abstract UI for the target platform (instead of going through the two transformations above), starting with the currently loaded (single-platform) task model, and using a number of default configuration settings related to the user interface generation.
- *From abstract user interface to concrete user interface for the specific platform.* This transformation starts with the abstract user interface and yields the related concrete user interface for the specific interaction platform selected. A number of parameters related to the customisation of the concrete user interface are made available to the designer. The tool generates XHTML and XHTML Mobile Profile according to the type of platform selected.

Evaluation Criteria and Methodology

Starting with the ISO 9241-11 standard definition [1] and Shneiderman's [3] and Nielsen's [2] metrics, but considering the double perspective of the tool itself versus the product realized through the tool, we identified four aspects to be evaluated and eight related requirements as listed in Table 1.

Table 1. Evaluation criteria

Aspect	*Requirement*
Tool Interface	Intuitiveness
	Learnability
Tool Functionalities	Completeness
	Developer satisfaction
Final Product Obtained with the Tool	User Satisfaction
	Maintainability and Portability
Approach Cost/Effectiveness	Development Efficiency
	Integrability

The experimental evaluation has been conducted in parallel to the tool development in order to provide a formative rather than a summative evaluation.

Two experiments have been designed in order to cover different aspects according to the criteria framework formerly exposed. Both of them refer to a common application scenario related to Business to Employee environment.

Five subjects, selected within Motorola GSG Italy staff, were involved in the evaluation. All of them, within a range of different background and specialization, have technical knowledge and experience in software design and development, and are experienced computer users. They have been asked to participate in a 30 minutes preparation session and to dedicate 10 minutes reading the TERESA help prior to start the exercises.

The first experiment focused on tool usability and functional coverage, with the objective to highlight potential weaknesses and to provide design recommendations useful while implementing subsequent versions of the TERESA tool.

The experiment consisted in starting with a given task model created with CTTE 1.5.7 and obtaining the concrete user interface for both desktop and mobile phone using the version 1.1 of TERESA tool. The exercise goal was to realize a simple version of an e-desk application allowing three main actions: the registration to the service by inserting a username and a password, the selection of a location (workplace, home, travel or vacation), and the selection of an application from a menu. The applications offered are different in the desktop and in the mobile versions of the service.

The actions to be performed, such as Generate Enabled Task Sets, Generate Abstract User Interface, etc. were predefined in order to require the access to every tool menu. For each step evaluators were asked to record any difficulties they may have encountered in achieving the goal and their suggestions to improve the user interface. In addition, they were invited to provide comments about: approach, functionalities and result produced, reporting advantages/disadvantages with respect to traditional methods and providing indications on additional functionalities they would like to introduce. The first evaluation provided a number of suggestions that were considered while developing the new versions of TERESA, such as the possibility of links between abstract interaction objects and the corresponding tasks in the task model so that designers can immediately identify their relations.

A second experiment has then been conducted in order to collect more information about developer satisfaction and cost/effectiveness of the approach. The experiment consisted in developing a prototype version of an e-Agenda application running on both desktop and mobile phone and including the following functionalities: visualization of the appointments of a single day; visualization of the details of each appointment; possibility of insert/modify/delete an appointment. This had to be realized in two ways:

1. At first using traditional techniques such as a template for the design phase and Microsoft Front Page or Netscape Composer for the implementation phase
2. Then using tool-supported techniques: CTTE 1.5.7 for task tree realization and version 1.5 of TERESA tool (updated taking into account the results of the first experiment) for XHTML and XHTML Mobile Profile pages generation.

The evaluators have been required to collect quantitative metrics related to development efficiency, such as the total effort needed to complete the exercise expressed as creation or rework time and categorized by process phase, as well as the number of errors introduced. Moreover, they have been required to express their judgment on specific TERESA characteristics such as support offered to individuate the most suitable interaction techniques, support offered to compose interactors in the interface, and others aspects related to developer satisfaction and product maintainability/portability by a rating from 1 (poor) to 5 (very good). In case of negative evaluation they were invited to provide an explanation note and suggestions for improvement.

Evaluation Results

The evaluation resulted in an amount of data about the aspects considered. As for the first experiment the analysis has been conducted in two steps. Firstly the raw comments have been abstracted to recurrent issues aggregated with a functional criterion, counting the occurrences of each issue; this step has been conducted iteratively, in order to progressively obtain a clean taxonomy. In the second step the taxonomy obtained has been presented to the evaluators, who were requested to express for each of them a relevance assessment, either high, medium, or low; from this new data a relevance index has been synthesized for each item. The results of analysis have been reported to the development group, which integrated them in the new version of TERESA used for the second experiment.

The new version 1.5 of TERESA has been substantially improved with respect to the first prototype. For example the effect of the heuristics used for combining together two or more PTS has been made more predictable, the AUI generation window has been redesigned in order to be intuitive and usable, the Final User Interface Generation has been improved by the introduction of a preview windows, the task corresponding to an object can be automatically identified, and some model editing options have been introduced.

The results of the second experiment show how developers' productivity is affected by the use of the tool. Data about time performance have been collected in each phase of the experiment and summarized through average values.

Results graphically illustrated in Fig. 1, show similar total times for the traditional and TERESA approaches, with opposite impact on different phases. The use of the tool almost doubled required time at design stage, while at development stage the results show a dramatically improved prototyping performance, reducing required time to half.

This leaves a margin for further improvement, since the design time required by TERESA approach is expected to decrease as the subjects become more familiar with model-based techniques and notations.

Moreover the slight total time increase is acceptable since it involves a trade-off with design overall quality: many subjects appreciated the benefits of a formal process and support to individuate the most suitable interaction techniques. For example designers reported satisfaction about how the tool supported the realization of a coherent page layout and identification of links between pages. The evaluators noticed and appreciated the improved structure of the presentations and more consistent look of the pages resulting from the model-based approach, as well as the reduced risk to forget the formal specifications. This is also coupled with an increased consistence between desktop and mobile version, pointed out by almost all the evaluators.

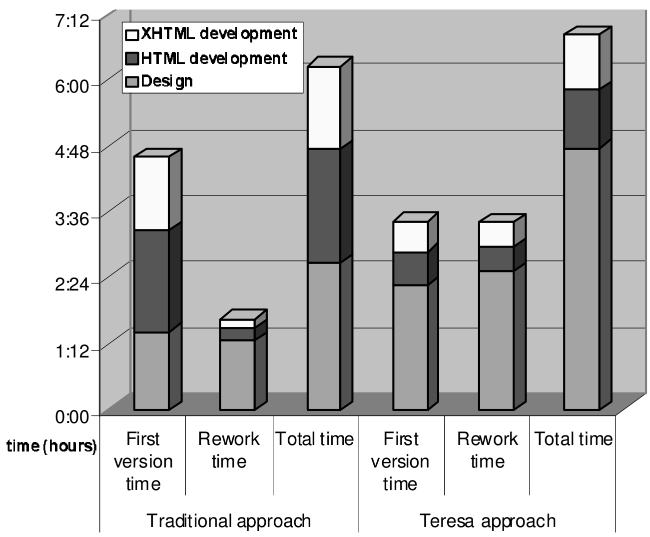

Fig. 1. Comparative results on time performance

Composition of time spent is also meaningful. The tool-supported methodology offers a very good support to fast prototyping, producing a first version of the interface in a significantly shorter time. On the other side rework time results increased.

According to evaluators' comments, root cause are the high level of automation of the tool, and the current unavailability of control options, which requires the developer to successively modify the interface produced in order to reach the desired results. Moreover, if changes and corrections are needed at the task model level, current prototype version of TERESA forces the user to a time-consuming iteration of the whole transformation process. The cost of corrections could be limited performing the platform filtering at a later stage of the process.

Another reason to take into account is the greater familiarity of the subjects with traditional techniques than with model-based techniques and notations.

Future refinements to TERESA and introduction of the tool in the software production process are then expected to consistently reduce rework time needed and to confirm the advantages of the proposed tool supported methodology.

Conclusions and Acknowledgements

In summary, TERESA emerged from the evaluation as an appealing and promising solution for designing and developing UIs on multiple and heterogeneous devices. Future refinements of the tool, are expected to consistently reduce the effort needed, and result in further improvement.

At the same time the evaluation methodology and criteria we introduced appears to be general and applicable to different systems.

Further activities will include additional experiments focusing on the final product and involving end users.

We gratefully acknowledge support from the European commission through the CAMELEON IST project. We also would like to thank the colleagues Cristina Barbero, Simone Martini, Bianca Russillo and Massimiliano Fliri for participating to the experimental evaluation and for the useful discussions.

References

[1] ISO9241-11 Ergonomic requirement for office works with VDT's – guidance on usability. Technical report, International Standard Organisation, 1991.

[2] Nielsen, J. *Usability Engineering,* Morgan Kauffman, San Francisco, 1994.

[3] Schneiderman, S. *Designing the User Interface. Strategies for Effective Human-Computer Interaction.* Addisson-Wessley, Reading (MA), third edition, 1998.

[4] Paternò, F., *Model-Based Design and Evaluation of Interactive Application.* Springer Verlag, ISBN 1-85233-155-0, 1999.