User Interface Evaluation : is it Ever Usable?

Christelle Farenc[a], Philippe Palanque[a,b][1], Jean Vanderdonckt[c]

[a] L.I.S., Université Toulouse I, Place Anatole France, F-31042 Toulouse Cedex, France, Tel: +33 - 61.63.35.{62,88}, E-mail:{farenc, palanque}@cict.fr

[b] Centre d'Etudes de la Navigation Aérienne, 7 av. E. Belin, F-31055 Toulouse Cedex, France.

[c] Institut d'Informatique, FUNDP Namur, rue Grandgagnage, 21, B-5000 Namur, Belgium, Tel: +32 - (0)81/72.49.75, E-mail:jvanderdonckt@info.fundp.ac.be

1. INTRODUCTION

According to human behavior studies, several disciplines (e.g., cognitive psychology, software ergonomics, visual design) have brought substantive results to improve the user friendliness of user interface (UI). One possible output of these disciplines come as recommendations that could be translated into ergonomic rules (or guidelines). Guideline knowledge is often contained in five sources : recommendation papers [1], design standards (e.g., ISO 9241 [2]), style guides which are specific to a particular environment (e.g., IBM Common User Access [3]), design guides (e.g., Scapin's guide [4], Vanderdonckt's guide [5]) and algorithms for ergonomic design (e.g., automatic selection of interaction objects [6]). Studies carried out with designers show that these guidelines are difficult to apply at design time:

- average search time for a guideline in a design guide lasts 15 minutes [1];
- about 58% of designers succeed to find guidelines relevant to their problem [1];
- designers do not respect about 11% of guidelines [7];
- designers experienced interpretation problems for 30% of guidelines [7].

Guidelines are intended to various people involved in a team which is responsible for developing an interactive application, but it is not clear how they best fit into their needs:

- a task analyst hopes that guidelines are organized according to a taxonomy of tasks characterized by attributes, and not according to interaction styles;
- a project leader only pays attention to high-level guidelines (e.g., selecting an appropriate metaphor) because they are the most likely to influence positively or negatively the success story of UI implementation;
- a human factors expert assumes that guidelines are sorted by cognitive principles that initiated the guidelines because they form the basis of work;
- a designer expects to see guidelines organized in such a way that s/he could instantly isolate specific guidelines and solve conflicts between them by ranking them into ordered sequences;
- a programmer thinks that guidelines should be organized with respect to the interaction objects (IO) or widgets of the interactive task because his/her responsibility only covers this part;
- a UI evaluator produces a UI evaluation report including scores, measures qualifying the degree to which guidelines have been respected, errors,...

2. UI DESIGN AND EVALUATION WITH ERGONOMIC RULES

Such a UI evaluation is performed not only in order to evaluate the UI quality, but also to improve the user friendliness of the product. Unfortunately, evaluation reports are not taken into account as expected because:

1. reports are not standardized : there is no unique evaluation method that could report all usability aspects and that could be teached as a whole. Instead, UI evaluators are forced to use multiple approaches which are not necessarily complementary and which are often grounded on human expertise and knowledge. The report structure depends therefore on both approaches and evaluators;
2. reports are not easily reused by designers and programmers : we just saw how the lack of standardization hinders a thorough reading and understanding of report by designers and programmers. Moreover, the report contents are not always presented to easily highlight usability problems in terms that programmers could manage. "Immediate feedback is not satisfied in 10% of cases", "Average task time is 3 min. 45 sec.", "Mean of error rate is one for fifteen transactions", "heart rate is 70 pulsations/min when achieving data retrieval, 90 when validating data" are examples of non-directly reusable conclusions.
3. reports are not prescriptive : if reports lists problems, the degree to which they are described may vary from one problem to another, from one evaluator to another, from one evaluation method to another. Moreover, it is the responsibility of designer to solve problems by finding relevant guidelines, successfull interaction techniques, to correct misleading dialogues,... And it is the responsibility of programmers to interpret these guidelines into practical IO programming.

2.1 Concepts

We suggest a knowledge organization that accomodates these differences among actors and that supports integrated use by both designers, programmers and evaluators. This organization consists of the concepts depicted in fig. 1.

Utility concerns the adequacy that should exist between functions provided by a UI and the user's task. *Usability* concerns the adequacy between the way an interactive task is carried out by a particular user and the cognitive profile of this user. When UI evaluation is performed, these two aspects should be handled, though usability is often emphasized. Utility and usability goals can be expressed as factors. A *factor* is a statement of a general evaluation dimension which is expressed symptomatically by intrinsic qualities and drawbacks and which could be measured and/or estimated. Several authors have their own factors :

- Shneiderman [8] enumerates five evaluation criteria: time to learn specific functions, speed of task performance, rate of errors, subjective user satisfaction, and human retention of commands over time;
- ISO 9241/10 standard [2] highlights several principles such as suitability for the task, self-descriptiveness, controllability, conformity,...
- Marshall, Nelson & Gardiner [9] call them "key categories, sensitive dimensions";
- Nielsen [10] argues that ten heuristics are sufficient such as: simple and natural dialogue, speak the user's language, minimize the user's memory load,...

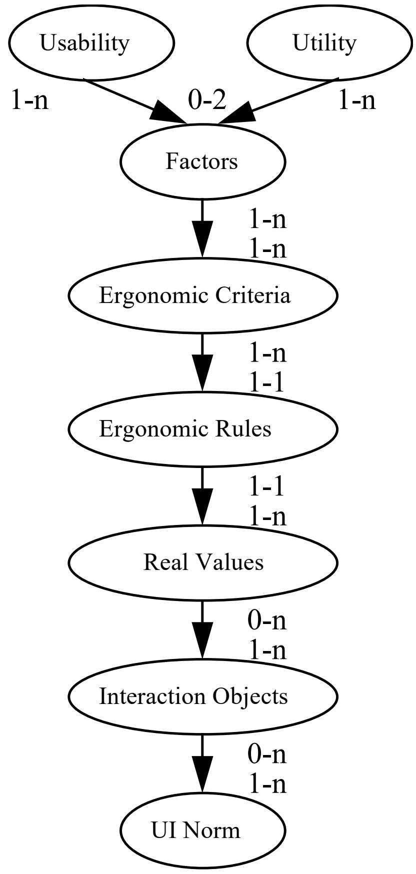

Figure 1. Structuring usability and utility into concepts.

A factor could be related to utility and/or usability (hence, the "0-2" connectivity in fig. 1), but, at least, one of them. Factors match user expectations for a usable UI and are related to one or many ergonomic criteria (hence, the "1-n" connectivity).

Ergonomic Criteria are widely recognized and accepted criteria that lead to an elaborated, efficient, sophisticated, user friendly UI [11]. They include for instance compatibility, consistency, work load, adaptability, dialog control, representativeness, guidance and error management as in [5]. Ergonomic criteria are primarily considered as design criteria because they can serve at design time, but they can also serve as evaluation criteria at evaluation time. For instance, it is interesting to see the impact of ergonomic criteria (e.g., consistency) on factors (e.g., rate of errors). If all ergonomic criteria might be considered for each UI, all criteria are not equally important: they should be weighted so that priviledged factors could be identified by examining important criteria respected by a UI. The satisfaction percentage of particular ergonomic criteria are a function of number and rate of ergonomic rules.

An *ergonomic rule* [5] consists of a design and/or evaluation principle to be observed to get and/or guarantee an ergonomic UI. This definition highlights that following guidelines is a necessary but insufficient condition to reach the goal. One can imagine that a same ergonomic rule can step in several ergonomic criteria (this is consequently a multi-criteria approach). But, if each recommendation from the literature is translated into a clean ergonomic rule with premisses and only one obvious conclusion, then each ergonomic rule realizes only one ergonomic criteria (hence, the "1-1" connectivity). However, it seems obvious that applying a particular ergonomic rule may influence the applicability and compliance of other.

The multiplicity of sources listed in the introduction inevitably mix empirical and conventional considerations. Premisses of some ergonomic rules may also depend on user stereotype parameters. We suppose that a particular ergonomic rule can be verified by comparing the real value of attributes coming from the IO and the ergonomic value praised by the ergonomic rule. For example, the "colour" attribute of a push button should be set to "grey" when this push button is inactive.

The *real values* are actually the observed values of all the IO composing a UI. The Look & Feel (i.e. the presentation and the behaviour) of common IO are partially regulated by norms (e.g., standards or style guides).

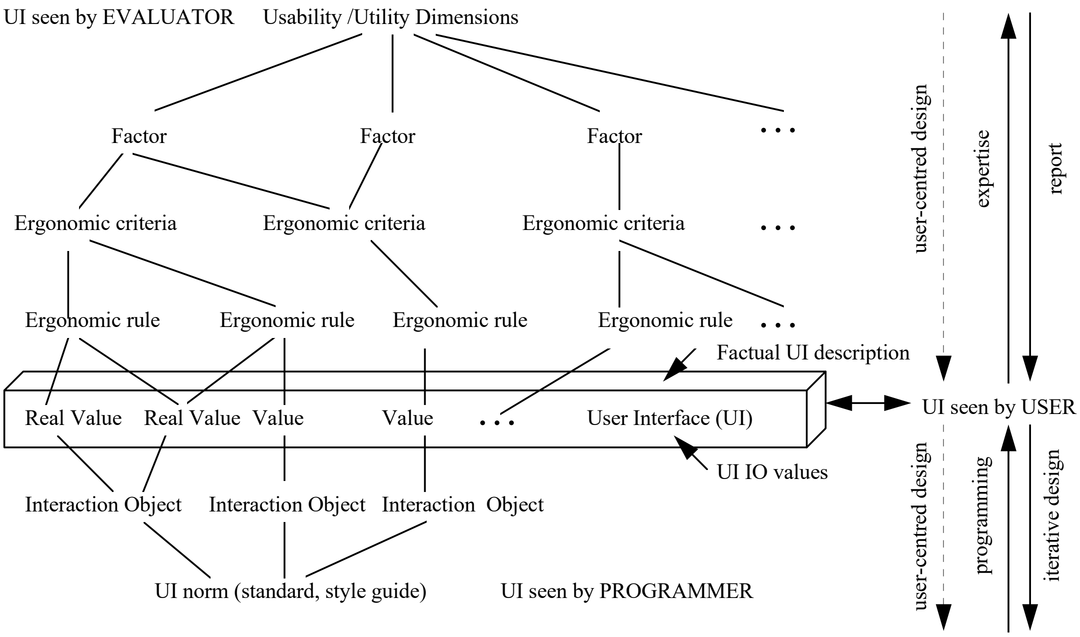

Figure 2. Multiple UI views.

2.2. Integrating UI views

Using these concepts aids integrating the multiple UI views possessed by three important actors (i.e. evaluator, programmer, and user) as depicted in fig. 2. *Evaluators* typically examine the UI of concern for finding usability problems. The results of this examination usually consist of strengths and weaknesses which are related to factors, which are themselves decomposed into ergonomic criteria which are in turn more related to UI properties and features. Partial or full respect of these criteria heavily depends on the ergonomic rules that are satisfied or not. Rather than expressing usability problems in terms of factors only, an evaluator could explicitly link them to ergonomic rules through ergonomic criteria. *Programmers* see a UI in terms of IO: whatever the development methodology is, UI design subsumes selecting and creating IO. If the evaluation report is expressed in terms of ergonomic rules working on effective IO, programmers are more likely to solve usability, utility problems. *Final users* perceive a UI by manipulating IO inducing easiness or difficulty in order to reach the task's goal. Highlighting the real values of current IO is a practical entry point for evaluating a UI with users.

Linking usability and utility to IO as introduced in fig. 1 allows us to use this approach for both UI design (dotted arrows in fig. 2) and repair (plain arrows in fig. 2). In the *first case*, one specifies which factors should be priviledged according to the UI context and user's demands. This enables us to compute appropriate values to be further satisfied for each criteria and to compare them with the real values. In the *second case*, we start from an existing UI ; we evaluate the compliance of ergonomic rules related to ergonomic criteria by computing the value for each criteria to be respected. This strategy embodies external parameters coming from context analysis (e.g., user stereotypes, socio-organizational context, work place). They are fundamental since we observe their real impact on ergonomic criteria, whether positively or negatively, allowing recognition of relevant factors.

3. Implementation Issues

In order to prove the feasibility of the above outlined integration of views, we examplify this by detailing two implementation issues. Tools for helping designer of interactive software have been implemented according to the above discussion. These tools try to embody as explictly as possible the organization defined in the previous section. These tools are grounded on a uniform knowledge structure which provides to the designer information compatible with knowledge domain

ERGOVAL is a tool which provides help for evaluating the UI usability. The system includes three sub-systems [12]:

- a rule base, containing ergonomic rules independent of the task and structured according to an interface object typology;
- a structural decomposition of the objects of the graphic interface;
- an interface object typology. Graphic objects concerned by the same recommendations were grouped into object classes. These classes formed a typology of objects with multiple levels of abstraction. Rules concerning an object class were implemented once, at the highest possible level, and are inherited by the graphic objects, which are the instances of object class.

According to this knowledge base, various organisations of ergonomic recommendations are allowed: design criteria-based (e.g., guidance, consistency), dialogue mode-based (e.g., command, action), graphic objet-based [3]. Moreover, the system reaches the following goals: usability by computer experts, time cost reduction when searching violated recommendations, completeness, cohesion, and maintainability of the knowledge base.

TRIDENT automatically generates a usable UI from context analysis (i.e. task analysis, definition of user stereotypes, and description of work place) with explicit use of ergonomic rules. In this system, a task-based approach is followed in order to conduct the presentation design. Ergonomic rules are not only imbedded and visible within the system, but are also documented and accessible on-line on request depending on the context. These rules could also be used to perform a UI evaluation by linguistic ergonomic criteria, though manual [13].

We do not claim that the knowledge organization outlined in fig. 2 still works for every evaluation method, but experience lead us to be convinced that it is practical for both surface evaluation (as in ERGOVAL) and by linguistic ergonomic criteria (as in TRIDENT). Not all UI aspects can be evaluated in terms of a comparison between a real value and an ergonomic value. For instance, it seems hard to evaluate if a particular metaphor is better than another since, even it is concretized as IO, the true reference values for appropriate IO materializing the metaphor remain unknown. Coupling automated UI generation and evaluation is considered as a key future work : it is sound to evaluate a particular UI that have been automatically generated with the same ergonomic criteria and guidelines that have been used to produce it. Moreover, ergonomic rules do not necessarily lead to an optimal UI because of conflicts and contradictions.

CONCLUSION

This paper has shown that the availability of guideline knowledge is not enough in order to provide helpful and efficient information to the agents involved in the design process of an interactive application. By showing the different points of view of those agents towards the UI we have explained why it is so difficult to provide relevant and easy to use information. Two different projects addressing this question are presented. The first one aims at evaluating and proposing solutions for the improvement of the interactive applications previously developped, while the second one aims at providing a set of tools for the automated design of interactive applications taking into account guidelines for user interface design. This paper has tried to give an answer at the question included in the title: without appropriate tools supporting both design and evaluation, it is no more allowed to follow the numerous guidelines currently available. Those tools are a necessary condition but of course not a sufficient one as the craft knowledge of user interface designers and ergonomists is far to be not worthy. User Interface evaluation: is it ever usable?

REFERENCES

1. S.L. Smith, Standards Versus Guidelines for Designing User Interface Software, in Handbook of Human-Computer Interaction, M. Helander (ed.), North-Holland, Amsterdam, 1988, pp. 877-889.
2. ISO/WD 9241, Ergonomic requirements for Office Work with Visual Displays Units, International Standard Organization, 1992.
3. IBM Common User Access Guidelines, Object-Oriented Interface Design, Document SC34-4399, IBM Corp. Publisher, 1993.
4. D.L. Scapin, Guide ergonomique de conception des interfaces homme-ordinateur, Research report INRIA N°77, INRIA, Le Chesnay, October 1986.
5. J. Vanderdonckt, Guide ergonomique des interfaces homme-machine, Presses Universitaires de Namur, Namur, 1994, ISBN 2-87037-189-6.
6. F. Bodart and J. Vanderdonckt, On the Problem of Selecting Interaction Objects, Proc. of HCI'94 (Glasgow, 23-26 August 1994), Cambridge, 1994, pp. 163-178.
7. F. de Souza and N. Bevan, The Use of Guidelines in Menu Interface Design : Evaluation of a Draft Standard, Proc. of INTERACT'90 (Cambridge, 27-31 August 1990), Elsevier Science Publishers, Amsterdam, 1990, pp. 435-440.
8. B. Shneiderman, Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison-Wesley, Reading, 1987
9. C. Marshall, C. Nelson and M.M. Gardiner, Design Guidelines, in Applying Cognitive Psychology to User-Interface Design, M.M. Gardiner and B. Christie (Eds.), John Wiley, Chichester, pp. 221-278.
10. J. Nielsen, Enhancing the Explanatory Power of Usability Heuristics, in Proc. of CHI'94 (Boston, 24-28 April 1994), ACM Press, New York, 1994., pp.152-158.
11. D.L. Scapin, Guidelines for User Interface Design: Knowledge Collection and Organization, Technical report ITHACA.INRIA.89.D12.03, INRIA, 30 Dec. 1989.
12. M.-F. Barthet, V. Liberati, M. Ponamale, ERGOVAL - A Software User Interface Evaluation Tool, in Proc. of the 12th Triennial Congress of the International Ergonomics Association (IEA) (Toronto, 15-19 August 1994) pp.429-431
13. F. Bodart and J. Vanderdonckt, Using Ergonomic Rules for User Interface Evaluation by Linguistic Ergonomic Criteria, in this proceedings.

[1]Ph. Palanque would like to thank the HCI Group of the Department of Computer Science of the University of York (U.K.) where he was a visiting researcher during the development of this paper.