Paper Prototyping - What is it good for? A Comparison of Paper- and Computer-based Low-fidelity Prototyping

Reinhard Sefelin, Manfred Tscheligi, Verena Giller
CURE Center for Usability Research & Engineering
Hauffgasse 3 - 5
1110 Vienna
Austria (Europe)
+43.1.743 54 51/11
cure@cure.at

ABSTRACT
This study investigated the differences between computer-based and paper-based low-fidelity prototypes. It researched whether subjects confronted with these two kinds of prototypes differ in their willingness to criticize a system and to give suggestions for its improvement. The chosen approach was an empirical study including test sessions using both kinds of prototypes. Quantitative and qualitative methods were applied to measure and to explain possible differences.

Keywords
Prototyping, Usability Testing, Design Methods

INTRODUCTION
Low-fidelity prototyping, as we understand it, is the visualization of design ideas at very early stages of the design process. The result is a prototype which is simple and whose development does not need very much time. A low-fidelity prototype can be vertical or horizontal. It may be developed using paper and other "low-fidelity-materials" or by the use of any user friendly programming tool.

This study investigated whether different ways of visualization and presentation of low fidelity prototypes affect the outcomes of usability evaluations in terms of subjects' critiques and suggestions to change a system. Although a lot of textbooks claim that paper prototyping enhances the communication between designers and users empirical data proving this assumption is not available. Also Bowling and Frick's [1] statement that a paper prototype makes users feel more comfortable criticizing the system was not proven by a comparative study. The available literature is full with success stories but empirical studies comparing different prototyping methods are rare.

From these view empirical comparative studies (see e.g. [4], [7], [3] and [6]) only two ([7] and [3]) compare two kinds of low fidelity prototypes using a user-based approach. Both of them come to the conclusion that the number of usability problems detected is not affected by the prototyping method, which is applied. However, the question whether subjects confronted with different kinds of prototypes differ in their willingness to criticize a system and to give suggestions for its improvement was not answered by these studies. This was the goal of our project.

CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.
ACM 1-58113-637-4/03/0004

APPROACH
Four prototypes of two systems were developed. One system was a calendar system, which enabled users to enter meetings, classes, birthdays and anniversaries. Furthermore, its user could request his/her entries in a daily and in a weekly overview. The second system was a touch screen ticket machine, which enabled users to buy tickets and to request information concerning their journey and concerning certain discount packages.

For both systems a computer-based prototype and a paper prototype were built. The functionalities of both types of prototypes were equal. The prototypes of both systems allowed the simulation of the system's most important functions. The paper prototypes were hand drawn. For a detailed description of the prototypes see [5].

24 test sessions and 48 usability tests were conducted. Every subject was confronted with one paper- and with one computer-based prototype of the two different systems. The subjects were divided into four groups. The order of the systems and of the type of prototypes differed between the four groups so that possible order effects could be avoided.

After an introduction phase, which was used to explain the purpose of the test session and to make the subjects feel comfortable, the subjects were asked to perform predefined tasks with both prototypes. For example we asked subjects to save a new entry in their calendars and to buy a ticket to Southampton. During the tasks subjects were encouraged to think aloud. After every task subjects had to rate the subjective difficulty of the task on a ten-point scale and they were asked to make suggestions to change the system in order to make the task easier. After the user had some time to play with the system he/she had to complete a questionnaire (SUS, [2]). Finally the facilitator asked the subject to summarize all the critiques and suggestions of

improvements that came to the subject's mind during his/her work with the system. During this phase the subjects were encourages to support their suggestions by hand drawn sketches. At the end of the second test a short qualitative interview was conducted concerning the advantages and drawbacks of the two prototyping methods. For a detailed description of the experiment see [5].

RESULTS

Table 1 shows the total number of suggestions to improve (change) the system and of the subjects' critiques. Note that every suggestion to change the system was also counted as a critique and that comments on graphical details were excluded. The analysis of the test sessions made sure that double counting could not occur. For a detailed description of the analysis of the test sessions see [5].

	Critiques			Suggestions		
	Mean	Std. Dev.	t-test	Mean	Std. Dev.	t-test
Ticket machine	8.96	5.49		6.96	4.33	
Computer prototype	9.25	6.96	t= -.25 p= .801	7.08	5.43	t= -.14 p= .891
Paper prototype	8.67	3.80		6.83	3.10	
Calendar system	10.17	5.40		7.83	5.48	
Computer prototype	9.25	4.14	t= -.82 p= .418	6.42	4.50	t= -1.28 p= .213
Paper prototype	11.08	6.49		9.25	6.18	

Table 1: Total number of critiques and suggestions

Table 1 shows that the number of critiques and suggestions is not affected by the kind of prototype. The table shows also the results of t-tests for two independent samples, which also did not show significant differences. A qualitative analysis dividing the subjects' statements into four categories (functions, operational design, behavior and screen layout, and wording) also could not uncover any major differences. A further qualitative analysis of the test sessions exposed two minor issues (neither of them was statistically significant): (1) subjects confronted with computer prototypes tend to comment more on graphical details. This trend does not lead to a smaller amount of other, more useful, comments. (2) subjects confronted with paper prototypes show a greater willingness to draw their suggestions. However, the qualitative analysis of the drawings showed that the suggestions, which were supported by these drawings, did not demand the usage of sketches to be understood by an experienced observer.

Also the analysis of the subjects' tasks ratings and of the SUS did not show any significant differences ($p > .2$).

During the qualitative interviews conducted at the end of each session 22 of the 24 subjects said that they prefer to work with a computer prototype. Most subjects stated that a computer prototype gives them more freedom to explore a system without causing "unnecessary" work for the facilitator, and that they feel "less observed".

CONCLUSIONS

This study showed two main results: (1) paper- and computer-based low-fidelity prototypes lead to almost the same quantity and quality of critical user statements and (2) subjects prefer computer prototypes. Since the comfort of subjects is one of the major factors of a successful usability test, one may argue that these two results mean that a design team should always prefer a computer-based prototype. However, there are still a lot of reasons to implement a paper prototype. The following list summarizes only three of them and is not meant to by exhaustive:

When should you possibly prefer a paper prototype:

- When the available prototyping tools do not support the components and ideas, which you want to implement.
- When you do not want to exclude members of the design team without sufficient software skills
- When the tests should lead to a lot of drawings, which then can be discussed inside the design team.

This list and our experience underlines that this study should not be interpreted as a rejection of paper prototyping. Our results just show that the decision on the appropriate prototyping method should also depend on the subjects' characteristics. If paper prototypes are faster and cheaper to develop still the problem of how to minimize the paper prototypes' disadvantages of giving subjects the feeling of being observed and of causing unnecessary effort for the facilitator will have to be solved. Future research in this area should focus on innovative methods on how to reduce these drawbacks.

ACKNOWLEDGEMENTS

We thank Claudia Kwapil for her help with the paper prototypes and Ann Blandford for her comments.

REFERENCES

1. Boling, E. and Frick, T.W. (1997). Holistic rapid prototyping for web design: Early usability testing is essential. In B. Khan (Ed), Web-Based Instruction. (pp. 319-328) Englewood Cliffs, N.J.: Educational Technology Publication, Inc.
2. Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In: Usability Evaluation in Industry. Taylor and Francis.
3. Catani, M.B. and Biers, D.W. (1998). Usability evaluation and prototype fidelity: Users and usability professionals. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 1998
4. Nielsen, J. (1990). Paper versus computer implementations as mockup scenarios for heuristic evaluation. Proceedings of IFIP INTERACT'90: Human-Computer Interaction. IFIP
5. Sefelin, R. (2002). Comparison of paper- and computer-based low-fidelity prototypes. CURE (Technical Report, available from the author)
6. Virzi, R.A., Sokolov, J.L., Karis, D. (1996). Usability problem identification using both low- and high-fidelity prototypes. Proceedings of CHI96, ACM Press
7. Wiklund, M., Thurrot, C., Dumas, J. (1992). Does the fidelity of software prototypes affect the perception of usability? Proceedings of the Human Factors Society 36th Annual Meeting, 1992