ELSEVIER

PII: S0003-6870(97)00027-6

Applied Ergonomics Vol 29, No. 1, pp 67–73 1998

0003-6870/98 $19.00 + 0.00

Testing usability with 3D paper prototypes—Case Halton system

Simo Säde,*[1] Marko Nieminen[2] and Sirpa Riihiaho[b]

[1]*University of Art and Design Helsinki, Hämeentie 135 C, 00560 Helsinki, Finland*
[2]*Helsinki University of Technology, Otakaari 1, 02150 Espoo, Finland*

(Received 31 July 1996; in revised form 21 December 1996)

In our study, we set out to see how low-fidelity three-dimensional paper prototypes could be used to test the usability of two alternative concepts for a drink can refund machine. The tests were carried out in the real environment with the actual users of the product. The tests took place before any software or hardware had been implemented. Using paper prototypes is a quantitative evaluation method which is easy to apply and suitable for product development processes with a tight schedule. We found the 3DPP modeling and testing method turned out to be useful and the results obtained had an influence on the product. © 1997 Elsevier Science Ltd

Keywords: Prototyping, user interface design, industrial design

Introduction

Halton System Oy is a Finnish manufacturer, and supplier of bottle and drink can refund machines and systems to shops and supermarkets worldwide. Since recycling glass bottles and aluminum material is nowadays seen as more and more important, the number of refunding machines is on the increase. At the beginning of 1996, Halton System Oy started a product development project to design a new aluminum drink can refund machine, which would be smaller and less expensive than the ones available on the market.

To begin with: a drink can refund machine basically works as follows: it takes the can from the shop customer, checks if it is an aluminum can and reads its bar code to see if it is refundable. Then it squeezes the can, throws it into a waste bag, counts the total number of the inserted cans and pays the customer either by printing a receipt to be cashed at the cashier or by supplying coins.

Nowadays, consumer products have to be developed at an increasingly fast rate, because of the need to be able to react to the changes on the market. In this case, due to changes in the legislation concerning recycling in Finland and Sweden, as well as on account of the situation on the market, a finished product had to be achieved in an exceptionally short time. Technical development was done by in-house and consulting engineers, and an industrial design consultant was hired to do the product design.

*Author to whom correspondence should be addressed

To ensure the usability of the emerging product, the company invited researchers from two universities (the University of Art and Design Helsinki and the Helsinki University of Technology) to be involved in the development work. This had already been done in the concept creation phase of the development cycle. At the universities, the researcher team works on a project called 'Smart Products—A Multidisciplinary Approach to User Interface Design', studying the usability issues of consumer and professional products. The group was interested in the ways of introducing usability to the product development cycle, in design rationale, and in prototyping industrial design as well as user interface design ideas. In this case, the usability experts focused only on the interaction between the end user and the product, whereas, for instance the shopkeeper's point of view was not looked into at this time.

Usability—An important criterion for a walk-up-and-use product

Usability has several dimensions. There is the well-known division into five measurable attributes: learnability, efficiency, memorability, errors and satisfaction (Nielsen, 1993).

A drink can refund machine is a typical walk-up-and-use product. When the user comes to the machine, no preparations are performed, no instructions are read, the user just starts to feed the machine with cans. The design of a refund machine might have as its goal the following: *all customers of a shop must be able to instantly use the machine*. Thus, an unambiguously designed product is called for.

The can refund machine being as simple a product as it is with regard to functionality, this goal should be easy enough to reach. If it is not, there is something wrong with the design. Using the machine must be so simple that once learned, it will subsequently be remembered. Simple machines are also likely to be fast and efficient, which contributes to the customer satisfaction.

Two possible concepts: automatic and manual

During the concept creation phase, many alternative solutions to a large number of open questions were discussed. These questions were interdependent, because a certain technical solution would lead to a certain industrial design solution, which in turn would have strong influence on usability, etc.

At a certain point, two different concepts were defined, although no final decisions had been made as yet. These two concepts were based on different technical solutions and their user interfaces were not alike. The feasibility of these alternatives had to be evaluated before the final commitment to either concept. The alternatives greatly differed with respect to usability because of the way in which tasks were allocated between the machine and the user in each of them. One of the alternatives was called the automatic and the other the manual one. The manual concept was supposed to be less expensive to produce. From the user's point of view, the main difference was how the cans were inserted into the machine (*Figure 1*).

The automatic concept was as follows: it had a rectangular opening in the front panel through which the can was placed on to a tray. The tray rotated, taking the can further into the machine, after which a cover closed the opening. After the machine had processed the can, the empty tray rotated back, uncovering the opening. Rejected cans would re-appear in the opening to be taken away by the user.

The manual version, on the other hand, had a round opening in its front panel. The user was supposed to push the can into the hole horizontally, with the bottom first and the bar code on the top. Otherwise, the machine could not read the bar code and would not pull the can in.

There was no textual display in either of the concepts, because the goal was to make a language-independent user interface. Both concepts had three indicator lights in their front panels.

- a red light accompanied by a 'remove-the-can' sign indicated that a can had been rejected
- a yellow light with a 'non-refundable' sign was lit for recyclable cans with no refund value
- a green light with a 'can-refunded' sign would appear for the approved cans

The yellow and green lights would be lit for a short moment while the can was being taken in. The red light would be on until the user would take the rejected can back. At the outset, the usability researchers suspected that the manual concept would be far too difficult to use. First, the users would not expect the location of the bar code to be of any relevance. Secondly, some users, such as children, are unaware of the existence of any bar code. Thirdly, the instructions and the feedback concerning the correct positioning of the can seemed ambiguous. For instance, if the can was put into the machine with the bar code facing down, the machine displayed the red light with a 'take-it-away' symbol, instead of displaying a symbol telling the user to turn the can around. If the can was put in with the top end first, the feedback was the same red light, although the user was not supposed to correct the mistake in the same way.

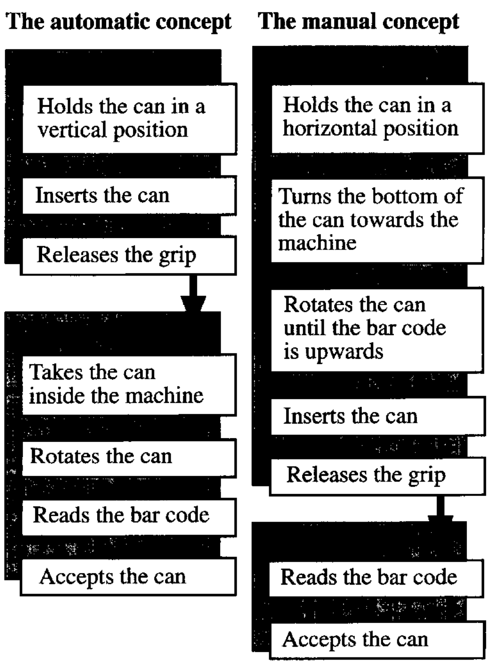

Figure 1 Flow-chart presentation of successful sequences of can insertion according to automatic and manual concepts

Modeling the concepts using three-dimensional paper prototyping (3DPP)

In order to find out whether the two concepts were usable or not, it was decided to test them with real users. The testing was to take place as early in the product development process as possible, before any hardware or software development. This approach was something new in Halton System's product development strategy. The Smart Product research team suggested that the test would be done using three-dimensional paper prototyping, a technique which was to be developed in the research project.

Three-dimensional paper prototyping (3DPP) is an application of the paper prototyping technique, which is used in the development of software user interfaces. Paper prototyping allows the prompt testing of design ideas in practice. The illustrations of a user interface are drawn on paper. These screen pictures are then presented to a user, who is asked to perform some representative tasks. The user presses the images of buttons just as he or she would do when using the real product. One of the designers simulates the feedback by changing the pictures in front of the user. This manual simulation is called the 'Wizard-of-Oz' technique. In this way, the user gets a hands-on experience of a nonexistent product, and design flaws can be found in the product concept (Casaday and Rainis, 1996; Kyng, 1995; Mayhew, 1995; Nielsen, 1993; Wilson and Rosenberg, 1988).

Two-dimensional images are of limited use when the size, form and relation of different components of physical three-dimensional products are being evaluated. We tried to apply the paper prototyping method to 3D by using fast-to-build cardboard mock-ups (*Figure 2*), similar to the ones that are used by industrial designers for sketching product ideas three

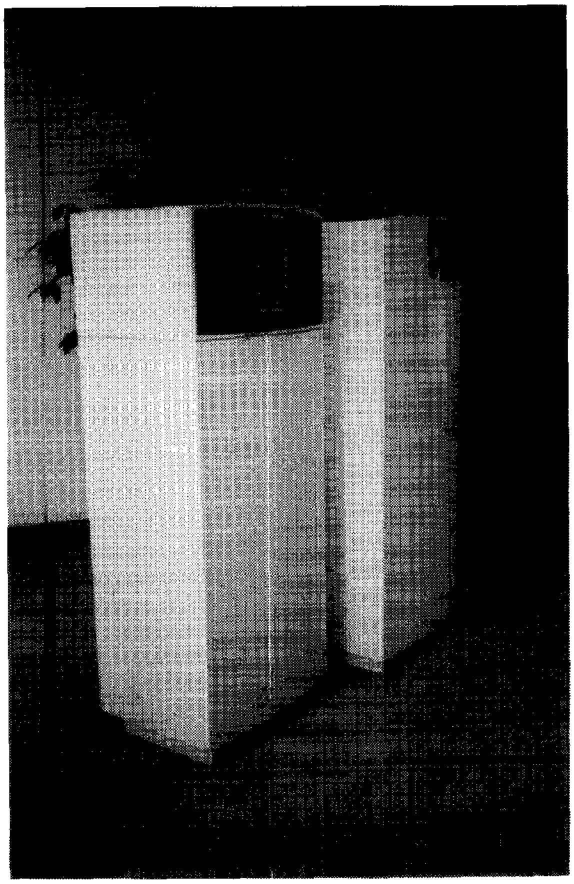

Figure 2 The automatic concept mock-up (left) and the manual concept mock-up (right)

dimensionally. This meant adding simulated functionality to these mock-ups, which is not usually done in the industrial design process.

The two mock-ups were cut out of foamcore cardboard sheets with a knife and the parts were glued together with a glue pistol. The mock-ups were full size. They had the white, smooth surface of the foamcore cardboard, with color prints of the front panel graphics glued on it. The indicator lights in the panel were correctly shaped dark spots. Turning on the lights was simulated with three pieces of colored paper that were attached to the panel with little pins at the right moment.

The Wizard-of-Oz was standing behind the mock-up with her hands inside it. In the automatic concept (see *Figure 3* and from behind *Figure 4*) the wizard turned the tray when a can had been placed on it. If the can was approved by the machine, the wizard took it away from the tray and turned the tray back empty. If it was not approved, the tray was returned with the can still on it. In the manual concept (*Figure 5*), the wizard performed the tasks of pulling the can into the machine when it was in the right position and stopping the can when it was in the wrong position.

Virzi *et al* (1996) define the level of fidelity of a prototype by four dimensions. The mock-ups in this case were quite low-fidelity both in terms of aesthetic refinement and functionality. Their appearance was not detailed nor finished. It was very obvious that they were only mock-ups. Their actual functionality level was non-existent, given the fact that they were only simulations. However, the similarity of interaction with the user and the breadth of features as compared with the real thing was very high. The user used the mock-ups in the same way as the final product would be used, that is, by placing cans into them and by getting feedback through the indicator lights. The number of features for the end user in this kind of product is quite low, and the only function that was left out of the mock-ups was printing the receipt.

Low-fidelity prototyping can be a valuable way of finding usability problems in the emerging products, especially at the early stages of the design process when there are many open questions and it is not reasonable to dedicate too much time on one problem. Low-

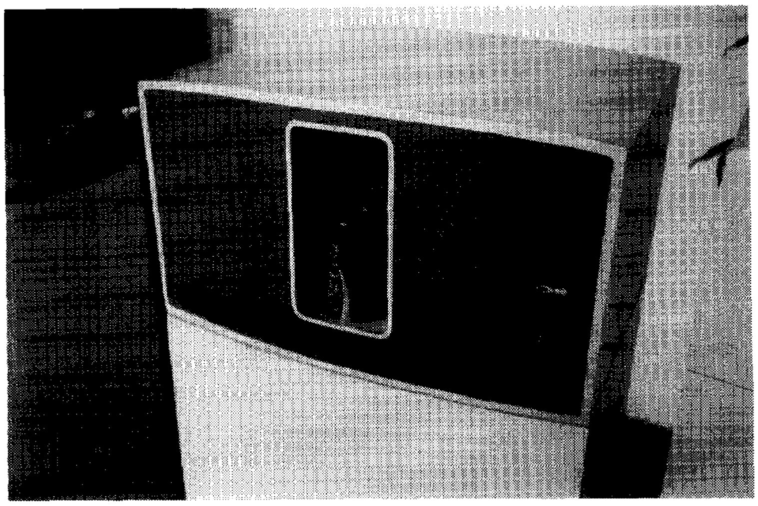

Figure 3 The automatic concept mock-up

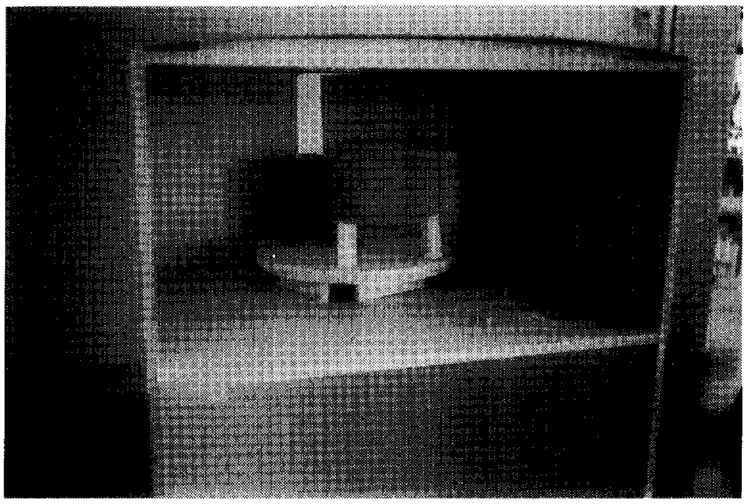

Figure 4 The automatic concept mock-up from behind

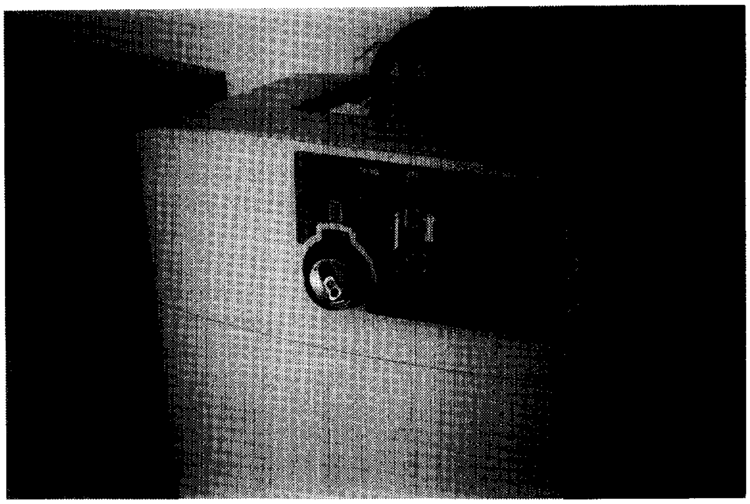

Figure 5 The manual concept mock-up

fidelity prototypes have proved to be an efficient way of detecting usability problems, but it is controversial whether the problems found with them are different from the ones found with high-fidelity prototypes (Nielsen, 1990; Virzi *et al*, 1996; Wiklund *et al*, 1992). Low-fidelity modeling cannot represent certain aesthetic aspects of design and consequently, it is unsuitable for situations where aesthetic aspects are of great importance, for instance, when the design solutions are to be approved by the management or client.

Testing the concepts with real users

The goals of modeling and testing of models were:

- to carry out preliminary tests on the usability of the product ideas with real users in a real environment already in the concept creation phase
- to compare the two alternative concepts
- to apply the paper prototyping method to work with three-dimensional mock-ups

An informal pilot test with two subjects was performed at the Helsinki University of Technology to find out if there were any problems with the test setting. After that, the actual test with 20 users was carried out in a large supermarket with a representative set of customers. This supermarket in Espoo town has customers from various social groups, aged from children to the elderly. Most of the shop customers in Finland are widely willing to take trouble to recycle bottles and cans because they are refundable. In the supermarket there was a recycling spot near the entrance of the supermarket with five bottle refund machines and one drink can refund machine. The two mock-ups were tested one at a time beside the machines. The three researchers were each assigned a different task during the test. Researcher A communicated with the supermarket customers and led the session. Researcher A also operated the faked indicator lights by sticking them on to the front panel (*Figure 6*) according to the dialog between the user and the mock-up. Researcher B acted as the Wizard-of-Oz (*Figure 7*). She was standing behind the mock-up facing the user, manually producing the simulated action. Researcher C made notes. The passing customers were asked to help in the test, and the reasons for the test were explained. It was emphasized that the purpose of the test was to evaluate the product

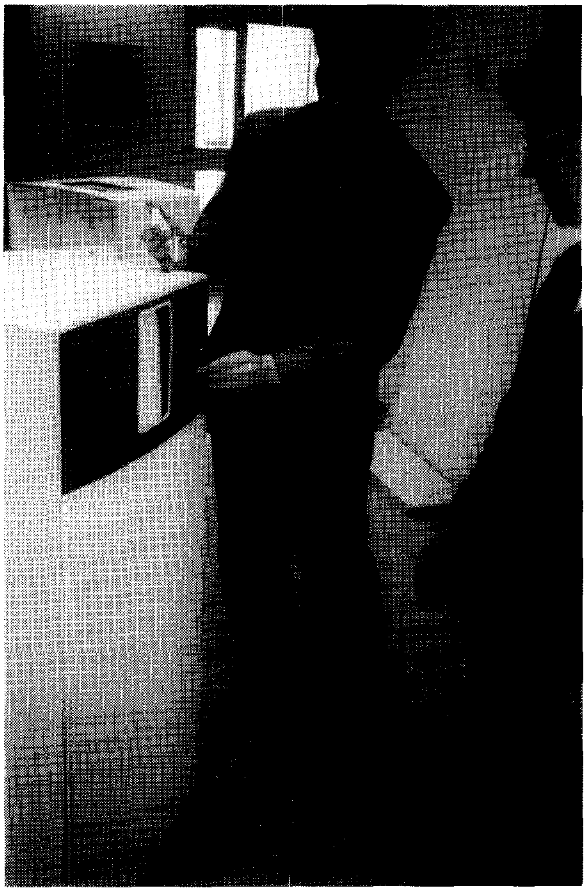

Figure 6 Interacting with the mock-up of the automatic concept, researcher A on the left, a 'user' on the right

Figure 7 The Wizard-of-Oz

idea, not the performance of the users, and that some problems with the prototypes were expected to arise. Voluntary subjects were given a plastic bag containing drink cans and they were asked to act as if they had come to recycle them and to put them into the machine. They were told that the colored pieces of paper were there to simulate the indicator lights. As the test did not require much time, no reward was given to the test subjects. Both concepts were tested with four female and six male subjects, aged from approximately 10 to 65. As the test was meant to serve only as a design tool, the number of subjects was considered sufficient. Virzi (1992) and Nielsen (1994) have shown that approximately 80% of usability problems can be detected using only five subjects. The gender distribution (40% female–60% male) is almost the same as in the results from the two observations of reverse vending machine users made in the same supermarket where the test took place (*Table 1*). The first observation took place at the supermarket after the mock-up test and the second observation when a functioning prototype had been built. The second observation will be described below, in 'Evaluating the Modeling and Testing Method'.

The results

The three main goals of the modeling and testing of the models were reached:

- users' ability to use the concepts was studied with real users in a real environment
- the two alternative concepts were compared and it was found that one was much easier to use than the other
- the three-dimensional paper prototyping method proved to be useful in this situation

The main result of this preliminary usability test was that the users found the manual concept clearly more difficult to use than the automatic concept. We must, however, point out that this test was performed to achieve an early approximation of the usability of the two product concepts, not to run a full-scale study with quantitative results. Anyway, what is clear is that the automatic concept proved to be feasible, unlike the manual one *Table 2*. On the basis of the researchers' observations, the user performance with the two 3DPP mock-ups was categorized as follows:

(1) Good: Users did not have any problems during the task.
(2) Minor problems: Users were occasionally confused, which slightly slowed down the performance.
(3) Major problems: Users encountered serious problems that slowed down the performance considerably. They were confused with the basic function of the concepts.
(4) Failed: Users were not able to return the cans.

The non-parametric χ^2-test was used to analyse whether the distributions of user performances with the two concepts are statistically different. The χ^2-test applied to previous results returns the probability value $p = 0.010339$. Thus, statistically significant ($p < 0.05$). About half of the subjects would not have been able to recycle their cans at all with the manual concept. The

Table 1 The gender distribution of the observed users of reverse vending machines

Sex	O1 N	%	O2 N	%	Total N	%
Female	7	25	17	34	24	30,8
Male	21	75	33	66	54	69,2

Table 2 User performance with the two 3DPP mock-ups

	Good	Minor problems	Major problems	Failed
Automatic	8	2	0	0
Manual	3	0	2	5

biggest problem was that many users did not realize that the position of the bar code was crucial. The subjects who used the manual mock-up also spent a longer time performing the task. This comparison between the manual and the automatic product concept is naturally dependent on the detailed design of each version. The manual one could be improved by iterative design. However, that would not change the fact that the users do not expect the position of the bar code to have any relevance, and that people normally do not check the instructions before using a device such as a refund machine. The test with 3DPP also revealed that only about one third of the users paid any attention to the simulated indicator lights. It was found that after placing a can into the machine, they usually looked at the bag containing the cans to pick up the next one. This took place simultaneously with the lights being lit. Only in the case of rejection did the subjects notice the red light more often. Although a real indicator light would catch more attention than a badge of colored paper, in paper prototyping, the hand that places new items on to the prototype also draws attention to itself. In this case, it is possible that it counterbalanced the difference between a real light and a paper light. In addition, the use of 3D paper prototypes revealed some other unexpected user actions that would not have been revealed using a non-physical 2D prototype only, such as placing the rejected cans on to the roof of the machine and not releasing the grip on the can when it was in the right position, etc.

The influence of the tests on the design

The tests with 3DPP mock-ups had a strong influence both on the design process and the decisions made. The usability consultants suggested that the automatic version should be chosen and that the indicator lights should be left out as they seemed to be unnecessary. These decisions were confirmed by the product manager of Halton System Oy.

The results of the preliminary usability test have also been utilized in marketing the product to possible clients during the development process. The number of products sold was already high even before the beginning of the production.

Evaluating the modeling and testing method

At the very early stages of the development cycle, the method can be used with the purpose of obtaining information for further development. Later, when the number of alternative ways to proceed has been reduced, more precise data will be needed. Then a prototype of higher fidelity and functionality must be built and tested.

The 3DPP technique is fast and cheap. In this case, the resources needed were three researchers working 5 h at the test session, and 25 h of work to build the mock-ups and to report the results. The material and tool costs were less than £100. This means that even several iterations could be made during the design process.

Three-dimensional paper prototyping suits certain types of products. It is capable of imitating the use of physical, three-dimensional interactive products. The less physical actions and objects there are related to the use situation, the less need there is for 3DPP. On the other hand, for clearly two-dimensional and virtual user interfaces, traditional paper prototyping with flat screen images would be an easier choice. In three-dimensional and traditional paper prototyping alike, very fast interaction and very large interface structures are hard to represent.

The preliminary usability test with 3DPP showed that people are able to forget the limitations of a low-fidelity prototype and concentrate on the task they are performing. In fact, the illusion worked so well that the subjects felt they really interacted with the mock-up. They talked to themselves saying things like 'why doesn't it take the can now?' and 'now the red light is on', etc. Even the members of the development team, who were familiar with the concept and the 3D paper prototype could feel the hands-on-experience when playing the part of the user (Ehn and Kyng, 1991).

Later, when the development work had progressed, a functioning prototype (*Figure 8*) was built at Halton System Oy. This was placed for a while in the same supermarket where the preliminary test had taken place.

The Smart Product researchers observed the users of the machine in order to find out if the 3DPP test had given an accurate estimation of the usability of the product concept. Another goal was naturally to detect usability problems which were not identified by the 3DPP test, such as ones concerning feedback times.

The concept had now become even simpler than before, since the indicator lights had been left out. The prototype was of quite high fidelity, and the customers took it as a real product.

Fifty users were observed. They did very well in the experiment. Of the 50 users, 41 had no problems whatsoever. Five users experienced some confusion that slowed down their performance. The prototype itself had transient technical problems with four users, who are excluded from *Table 3*.

Table 3 **User performance with the automatic 3DPP mock-up and the automatic protype**

	Good	Minor problems	Major problems	Failed
3DPP mock-up	8	2	0	0
Prototype	41	5	0	0

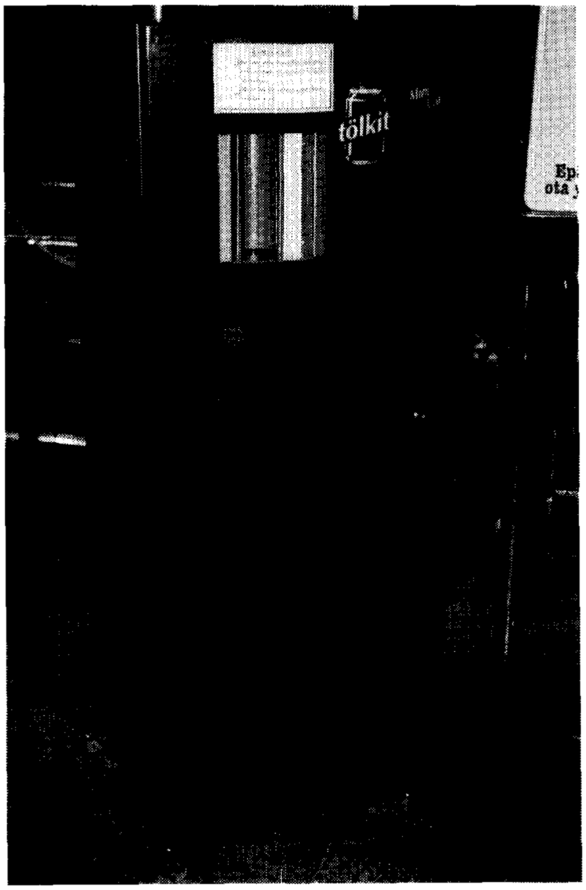

Figure 8 The functioning prototype

The probability (p) for the above results with the χ^2-test using the two categories (good and minor problems) is 0.429. Thus, there is no statistically significant difference between the results.

This suggests that in this case, the 3DPP mock-ups and user testing gave a correct estimation of the usability of the automatic product concept.

Conclusions

Three-dimensional paper prototyping can be a valuable development tool when estimations of the level of usability of early product concepts are needed. This knowledge can be gained before spending time and money on writing software or building functional prototypes. The mock-ups are fast and cheap to build, and can be tested with real users, who would not be able to comment on technical specifications or other abstractions, and who need tangible presentations.

3DPP prototypes are a combination of user interface design methods and industrial design methods. They represent the basic functionality, but also the basic design ideas in 3D. They allow the evaluation of the usability of the cognitive part of the interaction, the physical ergonomics, and the basic design ideas all at the same time. The relation of these three areas and the wholeness of the product concept are taken into account by representing the basic aspects of the product's appearance and use simultaneously.

The limitations of the method are its roughness and the fact that is is suited to only certain types of products.

References

Casaday, G. and Rainis, C. (1996) Requirements, models, and prototypes for HCI design. *Tutorial notes of CHI'96 Conference*, ACM

Ehn, P. and Kyng, M. (1991) Cardboard computers: Mocking-it-up or hands-on the future. In *Design at Work*, ed. J. Greenbaum and M. Kyng. Lawrence Erlbaum

Kyng, M. (1995) Making representations work. In *Communications of the ACM*, Vol 38, No 9

Mayhew, D. J. (1995) Managing the design of the user interface. *Tutorial notes of CHI'95 Conference*, ACM

Nielsen, J. (1990) Paper versus computer implementations as mockup scenarios for heuristic evaluation. *Proceedings of IFIP INTERACT'90: Human–Computer Interaction*

Nielsen, J. (1993) *Usability Engineering*. Academic Press

Nielsen, J. (1994) Estimating the number of subjects needed for a thinking aloud test. *Int. J. Human–Computer Studies*. Academic Press

Virzi, R. A. (1992) Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors* **34** (4), 457–468

Virzi, R. A., Sokolov, J. L. and Karis, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. *CHI'P96 Conference Proceedings*, ACM

Wiklund, M. E., Thurrott, C. and Dumas, J. S. (1992) Does the fidelity of software prototypes affect the perception of usability? *Proceedings of the Human Factors Society 36th Annual Meeting*. Human Factors Society, Santa Monica, CA

Wilson, J. and Rosenberg, D. (1988) Rapid prototyping for user interface design. In *Handbook of Human–Computer Interaction*, ed. M. Helander. Elsevier