Evaluating E-Commerce User Interfaces: Challenges and Lessons Learned

Rainer Blum and Karim Khakzar

Hochschule Fulda - University of Applied Sciences,
Marquardstr. 35, 36039 Fulda, Germany
{Rainer.Blum,Karim.Khakzar}@hs-fulda.de

Abstract. This paper presents lessons learned from a user interface evaluation - concerning the applied methodical approach. Four alternative implementations of product catalogue navigation and presentation components for online clothing shops were evaluated in a comparative study. The resulting rather complex experimentation setting revealed interesting issues for the design of similar experiments. After describing the study's methodical setup the paper analysis relevant aspects of the applied approach. Finally, lessons learned are derived that are of relevance for user interface testing methodology in related contexts.

Keywords: E-Commerce, User Interface Evaluation, Rapid Prototyping, Evaluation Methodology.

1 Introduction

Numerous scientific papers prove that not only product attributes and commercial terms and conditions are of high importance for the success of business-to-consumer e-commerce shops but also design-related characteristics of the shops themselves [1]. Concerning the latter, aspects like interactivity, usability, information presentation, media richness, accessibility, personalization, adaptation or experience-related traits (e.g. design quality, entertainment, playfulness etc.) have been identified as relevant and important. Therefore, e-commerce businesses are well advised to invest sufficient effort in an adequate consideration of these criteria when building and maintaining their online stores.

Developing e-commerce user interfaces for the B2C retail sector is a current subject in the research project SiMaKon, "Simulation of Made-to-measure and Ready-made Clothing Online for Fit Checking" at Fulda University of Applied Sciences. In doing so, the aforementioned design-related attributes are regarded as fundamental quality criteria. Concentrating on these issues, we early had to realize, that many fundamental user interface design options exist - not only for small details like user interface controls or structuring of individual pages, but even for the higher-level functional components of online shops, e.g. shopping cart or product catalogue. Due to the lack of standards or universally accepted guidelines on best practices for online shops, it was decided to conduct a controlled experiment comparing different design alternatives.

M. Kurosu (Ed.): Human Centered Design, HCII 2009, LNCS 5619, pp. 653–660, 2009.

Designing such a comparative experiment forces researchers to make challenging decisions: What shall, basically, be compared? Which attributes of the objects of comparison are suitable for a comparative consideration? How many objects can practically and reasonably be compared? Especially, when many alternatives exist - user interface design options in the case of this study - and shall be taken into account, experimentation settings may become rather complex.

These issues are, in an e-commerce setting, accompanied by other effecting factors: The audience is extensive and can become large and uncontrollable once an online shop is made public. The indispensable gathering of information about the targeted consumers is therefore not an easy task.

1.1 Objectives and Significance

Object of investigation of the research activities of the ongoing project SiMaKon is an innovative, interactive system for the support of apparel retail via Internet. The paper presents significant lessons learned in the context of a study on e-commerce user interface alternatives - concerning the chosen methodology.

As lessons learned in user interface testing are for a large part context-dependant respectively research-specific, details on the study's context are also provided. Thus, potentially, the reader is enabled to correctly interpret the findings and to transfer them to other research contexts. The results presented should help researchers working on similar problems in designing their experiments.

1.2 The SiMaKon System

Trying on and reliably appraising clothing in online shops with the help of individual virtual bodies constitutes the basic concept of the ongoing research project SiMaKon. Challenges addressed are:

- reliable fit checking and appraisal of optical characteristics in relation to one's own body respectively a virtual representation of it,
- seamless integration with product catalogue respectively product configurator,
- comprehensive interactivity, thorough usability and absorbing shopping respectively user experience,
- high customer acceptance.

Therefore, an integrated system, comprising of the functional components avatar generator, clothing pattern generator, clothing simulation, 3D rendering, e-commerce back-end and online shop front-end is developed and continuously evaluated. Some main practical implementation details are:

- virtual garment construction and sewing using real garment pattern data and realistic simulation of the garments' drape and the cloth surface,
- representation of subtle details, e.g. sewing threads and knobs, as well as high-quality lighting and shading,
- individual virtual human bodies (avatars) with configurable body dimensions, hair style, skin tone and face characteristics,

- rendered 3D content with avatar, complete clothing outfit and background scenery displayed in the Web browser, integrated in an online shop,
- customers can interactively change perspective and background scenery as well as zoom into details,
- pieces of clothing or made-to-measure garment components can be swapped directly from the product catalogue for easy comparison.

After entering the online shop customers may start by first configuring their avatar(s) or by browsing the product catalogue. There, interesting clothing can spontaneously be selected for try-on and appraisal with the currently activated avatar.

The user-related criteria is dealt with in a cyclic user-centred design process based on ISO 13407 [2], which involves rapid prototyping of design outlines. As, like already mentioned, many fundamental user interface design alternatives apply for the system's different functional components, comparative experiments with prototypic mock-ups are continuously conducted in order to come to well-founded design decisions. The functional component relevant for the study reported in this paper is the product catalogue navigation and presentation.

1.3 Modalities of Needs

One main aspect of the evaluation of the different product catalogue navigation and presentation concepts reported here was the notion of "modality of needs": In her work Michele Ambayé [3] describes three different ways e-commerce consumers usually proceed through the decision making process, depending on the consumers' current intentions. A consumer may start the shopping process with any one of the three types of needs.

Pre-knowledge driven customers know they have a need, what to look for (usually brand and product type) and where to find it on the Internet. The purchasing criteria are quite clearly defined in the consumer's mind. That means the consumers know what they are looking for. Thus, they search on limited criteria such as price and delivery. For example, a consumer has previously bought a pair of Levi's 501 jeans in a particular fabric and size online, and would like to purchase it again. When customers know the website, the price and the shop's service policy, they possess previous knowledge and act knowledge driven.

If only a vague idea of a need exists (product type), and the Internet is used for searching different e-retailers, the modality is called function driven. For example, a consumer has bought a pair of unbranded jeans before, but would like to purchase a different style now. Deciding to gather more information online, the consumer is looking for a product that fulfils a function (a pair of jeans for a particular purpose) rather than for a brand or style already well known to the customer. Thus, this mode applies if a consumer knows only what sort of product is needed, such as a suit shirt, but has not yet established specific criteria intrinsic to the product, such as for example its style or material. Therefore, the purchasing process is characterized by the need for the function provided by the product: For example, it has to be a shirt that is elegant and goes well with a suit.

Impulse driven consumers are just browsing the Internet and have no conscious knowledge of a need. Here, needs are established impulsively as a result of interacting with one or several websites. Browsing the Internet, a consumer is maybe looking for

some product or not really looking for anything in particular. A website or a sales discount advertising banner may catch their eye, showing a pair of Levi's jeans at a significantly reduced price. Even though the consumer does not really need a pair of jeans, the offer is too good to miss. The purchase decision is made impulsively.

It is important to note that people may possibly switch between modalities during the purchasing process (shifts in mode). This occurs, when the consumer's current need changes. For example, consumers may start with a particular need, but while browsing on the Internet for a particular product (function-driven), an advertising banner attracts their attention for a different product and they finally buy impulsively. Therefore, they began shopping in one mode and afterwards shifted to another mode.

For e-commerce user interfaces trying to optimally support modality of needs including shifts between them may be a promising way to best serve an otherwise extensive, heterogeneous audience. Furthermore, it is likely that a particular e-commerce user interface design may serve one modality of needs better than another.

2 Method

Four different concepts, each implemented with a rapid prototype with basic functionality, were chosen for the comparative study - based on three aspects:

- support of the depicted modalities of needs
- potential for high usability and user experience for a broad spectrum of apparel consumers and,
- innovativeness.

The four prototype concepts for the product catalogue presentation used in the study are "associative structuring", "automatic adaptation", "three-dimensionality" and "classic approach".

The associative navigation structure applied for the first prototype is aimed at releasing the user from having to understand and accept the usually firmly defined navigation structure predetermined by the designer, like e.g. an ordinary hierarchical structure. Instead, the sorting of the product catalogue is permanently adapted based on similarity among different product items, with and based on every activity of the consumer. The concept is described by Peake [4] and was transferred to an online shop context for the purpose of the study reported here.

Every click on a product image sets the chosen clothing item and its four attributes brand, colour, type and pattern as the new central topics of interest - and moves this product to the middle of the screen. All other items are immediately rearranged around it: The ones more similar to the centre product and, therefore, presumably most interesting to the consumer at this moment are moved nearer, less closely related items are placed more far away from the centre product. Each of the four considered attributes is dedicated a separate part of the screen space, with all products appearing in each of these distinct areas. Items that are not visible on the screen due to their distance from the centre space can be reached via scrolling arrows provided at the four edges of the screen.

Automatic adaptation of the interface to the current modality of needs constitutes the main aspect of the second approach. Part of the navigation of the product

catalogue is changed according to the customers' current mode. For this prototype a hierarchically organized product catalogue was complemented by additional navigation aids. These are

- an advanced search form with search filters for product features like colour, material or pattern for pre-knowledge driven customers,
- a navigation structure with all featured trademarks as root categories and types of garments as subcategories for function driven mode,
- a navigation structure listing and sorting the products under different thematic, lifestyle oriented aspects like winter, sports or outdoor for impulse driven consumers.

The third prototype mimics the three-dimensionality of a real store, allowing customers to explore the store's range of products similarly to real life. Separate product items are displayed in two-dimensional information spaces each located on a separate wall section. Using several virtual rooms on several floors the product range is structured spatially. Outline maps in the form of a schematic presentation of the room and level structure allow for the transfer between the separate spaces. Other navigation means, all used via mouse and keyboard, are moving forward, backward, to the right and to the left, turning around at the current position and automatically moving along a predefined path to a certain destination, while observing the crossed environment.

The fourth and last prototype concept is a classic, hierarchical presentation of the product catalogue. It was deployed for comparative purposes, to see if one of the other concepts has advantages over this traditional way to present a product catalogue. It consists of a hierarchical tree displaying the product categories on the left and a detail view in the middle. The detail view presents either the items belonging to the currently selected product category or detailed information about a chosen, single product.

Sixteen mixed-gender candidates were tested individually in laboratory conditions, one person at a time in a one hour session, supported by two test assistants. Approximately 80% of the test persons belonged to the age groups 16-24 and 25-35, with gender uniformly distributed and five persons never or infrequently (less often than every two months) buying products via the Internet.

The study employed a three (three modalities of needs) by four (four prototype concepts) factorial design. Both factors were manipulated within subjects. Modality of needs was controlled by scenario descriptions that were presented to the participants, each detailing a typical situation for the respective modality. For example, for the function driven modality it stated: "Imagine, you are currently intending to buy several pieces of clothing with quite specific characteristics, e.g. a certain type of clothing, a certain colour, a certain pattern or several attributes combined. Therefore, you are browsing the catalogue of an Internet shop for products having exactly these features". Prepared like this, the test subjects had to accomplish several given tasks with each of the prototypes. The tasks were characteristic for the current modality, e.g. "Please choose a men's shirt, basically of the colour blue with a checked pattern." After having passed through all the modality scenarios with a prototype, the next one was tested.

During task execution the test assistants collected usability-related performance measures, namely success, speed and effort of completion, occurrence and severity of difficulties. Furthermore, they noted observations and estimated user experience

attributes on five-point Likert scales, with the items satisfaction, excitement, interest, joy and fun.

After completing a condition, the participants were asked to rate the tasks' complexity. Then, for each prototype, general usability-related questions were raised, involving estimation, e.g. usefulness for the given task, and open-end questions, e.g. if the subject especially likes a specific feature of this design alternative and why. This was followed by prototype-specific questions, e.g. concerning characteristic features. Finally, the participants had to appraise the same user experience attributes as mentioned above.

3 Results

Concerning this section the reader has to be aware, that it does not provide a description of the results of the described prototype study itself, but of relevant issues arisen from the methodical setup of the experiment.

The first aspect concerns the scenario descriptions presented to the participants. Each detailing a typical situation for the respective modality, they were targeted on controlling the modality of needs currently effective. Indeed, it may be questioned if "putting" a person "into" a specific modality of needs with a scenario description actually has the desired effect. No conflicting evidence occurred in the study, but, in fact, this issue was not well examined.

Then, the rapid prototypes may have been conceptually and functionally too different for a meaningful comparison. For example, the associative prototype was based on constant resorting of the product catalogue, while the 3D prototype did not offer sorting or filtering at all. Likewise, the adaptive prototype was the only alternative that "reacted" especially to the current modality.

Also, the maturity of certain aspects of the rapid prototypes may have varied to much. This concerns for example differently developed degrees of self-descriptiveness of separate interaction elements, e.g. readily identifiable scroll arrows for the associative prototype versus an abstract navigation map in case of the 3D prototype. Analysis of results also manifested that the applied principles were not clear or easily perceivable for every alternative. For example a significant number of participants did not comprehend the sorting principle of the associative prototype. Admittedly, this may put the results achieved into question. The trade-off between affordable effort and required quality proved to be an essential problem in rapid prototyping for this user interface evaluation.

Furthermore, one salient problematic aspect proved to be the prompted and observed user experience attributes. The statements and self-evaluations of the participants often differed considerably from the test assistants' notes. Clearly, definite tendencies were identified for many points and are presumably stable. Nevertheless, a significant amount of uncertainty concerning the results' validity in this domain remains, notably because neither a positive nor a negative tendency was identified by the test assistants for a great part of the participants. Insufficient qualification of test assistants or lack of quality of the observation criteria may be responsible factors. Also, it is worth discussing if these aspects may be reasonably observed at all. At

least, to reduce subjective differences, the same test assistants may better have carried out all the tests. However, it is unclear if crew change really has affected the results.

Another interesting aspect is that test assistants suspected that the mood of the test candidates or their physical and mental state on the day of testing may have had a significant effect on the achieved results.

Concerning the formation of the test participants group, it should be cared for more heterogeneity: More persons of higher age groups and of different education strata should be included for the present context, online shopping of clothing. Furthermore, only those may be considered that actually buy clothes via the Internet, for the sake of significant results.

Positively, scale and one hour duration of the test seemed to be well-chosen. Conducting part of the interview after each prototype was effective as well, as the participants' impressions were still fresh and not diffused by the interaction with the alternatives.

4 Conclusions

From this work in an B2C e-commerce setting, the following lessons learned can be derived for the methodical design of user interface tests with rapid prototypes.

First of all, be sure that scenario descriptions aiming to transfer test candidates into different, mutually exclusive states respectively situations really serve their purpose. At least, try to check if the scenarios are actually "adopted" by the participants, for example via suitable verifying questions.

Ensure that in the case of comparative experiments, different test situations respectively prototypes are not too varying for a meaningful comparison. Better concentrate on one aspect of comparison. Instead, several tests on differing characteristics of a smaller extent should be conducted, though this may often be tedious. Ideally, only a small number of objects should be compared in one study in order to limit the complexity of the experimentation setting. In the present case, comparing two different versions of a 3D shop environment may have been more effective.

Then, it is important to ensure comparable maturity and functionality for all relevant aspects of the included rapid prototypes. This way, unintentional bias on results can be avoided. Furthermore, design evaluation criteria for prompted and estimated user experience attributes thoroughly. Double-check mechanisms should be applied to ensure validity of results.

Acknowledgments. This research was financially supported by the German Federal Ministry of Education and Research within the framework of the program "Forschung an Fachhochschulen mit Unternehmen (FHprofUnd)".

References

1. Chang, M., Cheung, W., Lai, V.: Literature derived reference models for the adoption of online shopping. Information & Management 42, 543–559 (2005)
2. ISO 13407, Human-centred design processes for interactive systems (1999)

3. Ambayé, M.: A Consumer Decision Process Model For The Internet. Brunel University Information Systems and Computing PhD Theses (2005)
4. Peake, R.: An Experiment in Associative Navigation (2007), http://www.robertpeake.com/archives/365-An-Experiment-In-Associative-Navigation.html