A Comparative Usability Evaluation of User Interfaces for Online Product Catalogs

Ewa Callahan
School of Library and Information Science
Indiana University
Main Library 001
Bloomington IN, 47408 USA
+1 812 855-6268
ecallaha@indiana.edu

Jürgen Koenemann
humanIT
Human Information Technologies GmbH
Rathausallee10
D-53757 Sankt Augustin, Germany
+49 2441 92926-66
koenemann@humanit.com

ABSTRACT
We conducted a comparative study of the usability of hierarchically structured and zoomable Electronic Product Catalogs (EPC). After evaluating 16 hierarchical EPC available online today we constructed a test search interface and used it for comparison with InfoZoom, a novel interface utilizing a compression technique, direct manipulation, and zooming. 26 users, who were new to the InfoZoom Interface, but familiar with the Web environment, were studied in their use of one of these catalog types. Performance on both types of interfaces was assessed in terms of search speed, accuracy, efficiency, and user satisfaction. Both interfaces had positive evaluations, but behavioral data and performance data suggest that users valued their interaction with InfoZoom more.

Keywords
Human-Computer Interaction, user interfaces, usability evaluation, electronic catalogs, e-commerce.

1. INTRODUCTION
The rapid development of the Internet-based World-Wide Web (WWW) and its constantly growing popularity have motivated many companies to use it as a medium to promote and sell their products and services. The increase in electronic commerce creates a need for Electronic Product Catalogs. Benefits of online catalogs for manufacturers are obvious: lower distribution costs, the possibility to include more detailed information about the product, frequent updating etc. The customer enjoys a higher level of interactivity such as a variety of searching possibilities, comparison features to get access to more information about different products etc. However, navigation problems are (after security concerns) the second most cited reason for not shopping online (see e.g. [7]).

There are not many published studies to date on the *usability* of EPCs, but the number of publications about the general topic indicates a great interest in this area. Many of these publications emphasize the limitations of existing approaches and suggest guidelines for developers. Some attempt to transfer existing research on marketing and consumer decision making to the new application area.

Maes, Guttman and Moukas [10] developed a model based on traditional consumer buying behavior research. They identified six fundamental stages of the buying process: (1) need identification, (2) product brokering, (3) merchant brokering, (4) negotiation, (5) purchase and delivery, and (6) product service and evaluation. EPC's facilitate the first three stages.

To better understand the use of EPC's one should take consumer goals and tasks into account that mediate the searching for a product. Bettman and Kakkar [3] identified two ways of choosing a product:

- Choice by Processing Brands – a customer looks at particular brand and gathers information about some of its attributes, next the customer may examine a second brand, gather information on attributes (not necessarily the same like for the first one) and so on.
- Choice by Processing Attributes – a customer acquires information about products by looking first for particular attributes and examines its values for different brands and next moves to another attribute and repeats the process.

The choice of the process used is affected by the structure of the tasks [11], the way the information is displayed [3], and the attribute's value type (verbal or numeric) [12]. Consumers seem to process information in the way that is afforded by the information display. For this reason, if some methods of processing information are easier or more effective for consumers than other, information should be presented in a format congruent with those methods of processing. Displays should also take into account various consumer tasks.

A well-organized interface can have a significant influence on the store's profits. A study by Lohse and Spiller [9] suggests that improving the browsing and navigation capabilities of the e-store interface allowing the consumer to compare products can increase traffic and sales. Carefully designed product lists with additional information like product description and pictures also have a great impact on sales.

EC'00, October 17-20, 2000, Minneapolis, Minnesota.

There are many different catalog interfaces present on the Web. The most common are hierarchically organized catalogs, where the consumer sees the list of the products (usually organized in categories) and can follow the links to their more detailed descriptions. As easy as it may seem to use, these interfaces tend to be confusing for the consumers. The navigation of some sites is poorly designed. Items in dropdown menus are not visible. Lengthy result pages and lack of filtering options force consumers to scroll through long pages to find desired items [16].

There are two general solution approaches to solve the interaction problems in EPCs: increased automation and increased and improved interaction. The first approach for example includes recommendation systems which suggest to the user what the user should buy (recommendation approach). This approach is often favored by marketing people. It either requires the user to enter personal data prior to shopping in order to find the desired product which many users will not do (cf. paradox of the active user)[4]. Other solutions in this direction are based on gathering information from various customers or finding similar products, and creating recommendations on that basis[13](collaborative filtering approach).

The second approach concentrates on improving and facilitating the consumer's interaction with the interface, allowing the user to select products more easily (interaction approach). We were interested in this direction and wanted to investigate whether or not highly interactive interfaces that support the user in the select ion of products can improve the finding of products (and information about these products). One of the promising tools from this category is InfoZoom[14] which was used in this study.

2. EVALUATION OF EXISTING ONLINE CATALOGS

To evaluate InfoZoom we decided to compare it to a hierarchical type EPC commonly used on the Web. As a first step we examined various EPCs, paying close attention to how they were organized. For our evaluation, we selected 16 German and English automobile EPCs available on the Web in the Fall of 1999. Two sites were later eliminated from the evaluation due to serious technical and navigational problems (for a complete list of sites see Appendix). Our choice of automobiles as shopping domain was due to a number of reasons: car catalogs are very popular on the Internet, they cover a well-known and interesting domain for many end-users, and cars are objects that can be described by many attributes of different types (images, text, numeric, Boolean).

After an initial screening of selected sites we developed a feature space for evaluation and comparison and examined each of the sites according to a set of criteria we developed. These criteria were driven by our understanding of a typical product search for a more complex product such as an automobile as an iterative process of satisfying ones dynamic information needs [2]. We were interested in selection features offered by these sites and in interface design solutions for supporting an effective search. We checked how values of attributes are chosen (free text entry or selection), the possibility of specifying ranges for numeric values, and if multiple values for attributes are allowed. Additionally, we paid attention to support for iterative search steps, e.g. through a search history. We also recorded how search results are presented (on a single page or broken up into many smaller pages), if and how results can be sorted and resorted and if items can be compared in detail. Table 1 depicts a summary of the results.

2.1 Database size

Some sites lacked information about the number of cars in their database. This makes the search more difficult, especially following an attribute mode of searching. The consumers have trouble estimating the number of criteria needed to be chosen in order to get a reasonable number of hits. Additionally, in small

Table 1: Comparison of hierarchical interfaces used in commercial B2C car catalogs

Site	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Database														
Type of Cars	Both	Both	Both	USED	USED	Both	Both	Both	Both	USED	New	USED	USED	Both
Size of Database	?	?	460	?	?	250000	?	?	?	?	1433	N/A	23422	?
Search														
Free search	No	No	No	No	No	No	No	Yes	No	Yes	No	No	No	No
History	No	No	No	Some	No	Some	No	No	No	No	Some	No	Yes	Yes
Entry or Selection	S	S	Both	S	S	Both	Both	Both	Both	Both	S	Both	Both	S
Multiple Values	Yes	No	No	No	No	Yes	Yes	No	No	No	Some	Yes	No	Some
Search Results														
Number of Hits	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Sorting	Yes	No	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes
Resorting	Yes	N/A	N/A	Yes	N/A	Yes	N/A	Yes	N/A	Yes	N/A	No	No	No
Result Pages	G/20	1	G/20	1	G/10	G/50	G/5	1	1	1	1	G/10	G/20	1
Comparison	Yes	Some	Some	Some	Some	Some	Yes	Some	Some	Some	Yes	Some	Yes	Yes
Nr. of Attributes	24	6	7	8	7	9	41	5	10	7	80+	5	5	9

databases, in cases where no car fulfilled the criteria, adjusting the attribute's value is difficult. In some cases, the exact number of available objects can be obtained from information given on the search result screens for searches without restrictions.

2.2 Searching mode

Examined catalogs offered two ways of searching for a car. In the case where the customer had a particular car in mind and wanted to gather information whether or not this car was available, the user was referred to a page that allowed searching by simply choosing the make and car model (we label this type of search *Object mode*). This search was more common for new car searches rather than used car searches but there were exceptions. Customers could also search for a car using criteria other than make and model (such as price, acceleration, and engine type) (*Attribute mode*). Some sites offered combined searches (make and model were among the attributes). Two of the sites offered a free string search.

2.3 Attributes

In attribute search mode, attributes were combined by Boolean AND. In case a customer did not choose a value for some attributes it was considered as "all" (model, type) or "don't care" (price, mileage).

In many cases, it was not possible to search for more than one value for an attribute (e.g two brands) at a time. It was typically possible, however, to search for all values at once (options "all" or "don't care"). In a great number of cases, values had to be selected from pull-down menus and check boxes. Exceptions were made sometimes to attributes such as car model and price range that could be specified by entering free text into a text box.

Sometimes, choosing cars from static pull down menus was not the best solution. Sites for used cars often listed choices that were not (no longer) available or allowed users to combine attributes in way that –in general- had no matching item. Some impossible combinations are annoying but obvious to users –e.g. if car models are listed that do not belong to an already selected brand – whereas many others are difficult to discover for the end-user. The customer has to go through all the steps of selection to find out that the particular model or feature set is not in the database at that time. The underlying reason is often the coarse grain-size of the interaction between the client browser and the server: the user has to select multiple attributes on a single page and submit these selections explicitely to the server (search button etc.) before the next page loaded is reflecting choices made up to the time of submission. Most sites had dynamic adjustments of the interface option after selecting some criteria (usually make): after a make of a car was chosen the customer was offered only models (or car types) that were associated with this make.

2.4 Specifying Ranges

Ranges for numeric values of attributes (price, mileage) were usually assigned by site designers and could be chosen by a customer from the pull-down menus. Most ranges were not designed correctly. In some cases it was not clear if the selection

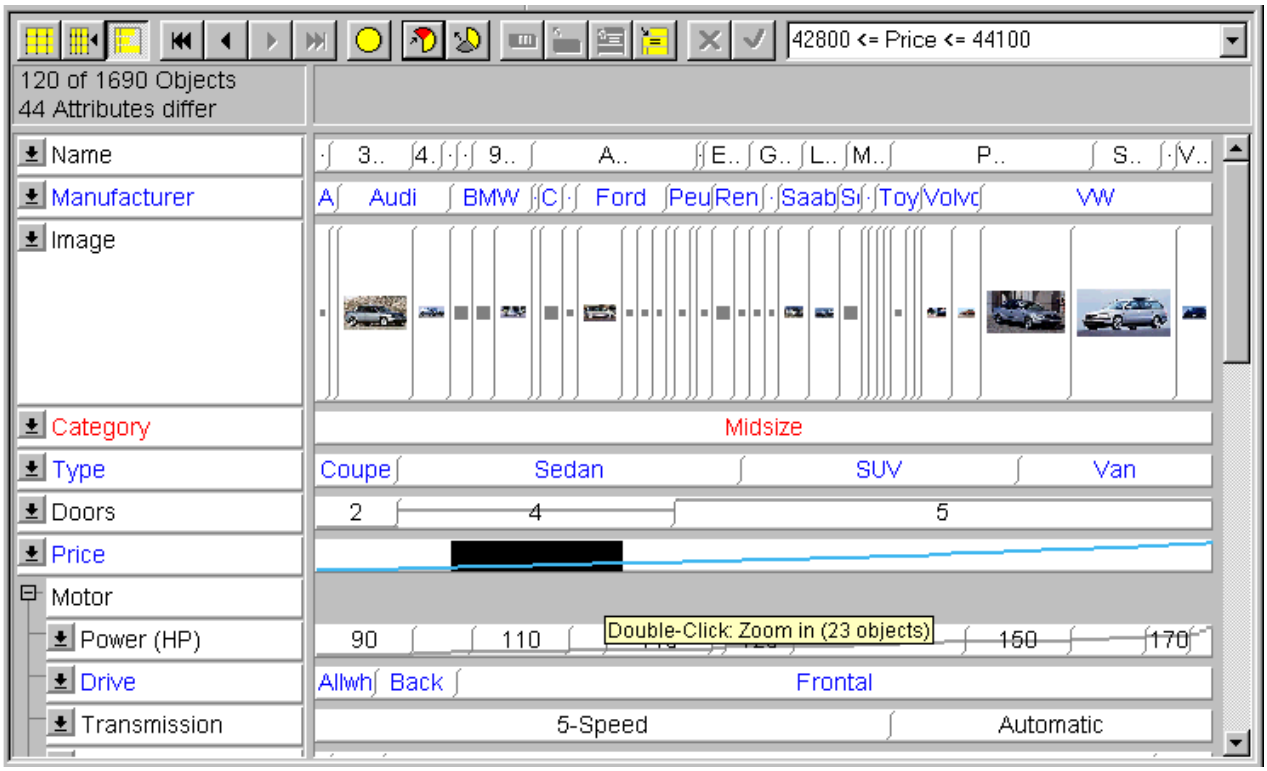

Figure 1: InfoZoom – Overview Mode

value represented maxima or a specific number (e.g. “4 doors”). In particular, static ranges presume that users have known common thresholds or have thresholds that are rigid.

2.5 Search History

Many sites did not provide any search history (in some cases it was understandable due to the simple search structure) other than the navigation offered by the web browser (back). Some sites offered information on how many cars are selected from the database after each step of the search and a return to the previous search.

2.6 Result Pages

The structure of the result pages was similar in almost all sites. The results were presented on one page or grouped and presented on consecutive pages, where they could be compared on the basis of limited, designer-set criteria. For each car found a link to a detailed page followed.

2.7 Sorting Search Results

The specification of the sorting criterion was usually possible on the search page through selection from a set of given options (e.g. by make or by price). However, resorting was only possible by going one page back and changing the criteria. Only in one case was resorting possible on the result page. In case a search brought no results, many sites did not provide detailed information on why this may have happened. The customer knew that a car with chosen criteria was not in the database, but did not know which criteria should be adjusted to find a car. It was especially difficult in databases of used cars with small numbers of offers at a particular time. Additionally, due to the poorly designed search structure the customer does not know if the car is not available at that given time or chosen criteria are mutually exclusive by design.

2.8 Item Comparison

All catalogs offered some kind of car comparison on the result page, usually by 4–9 attributes that were pre-selected by site designers. Car comparison as a separate option was offered only on three sites. It was offered as a simple search (Find two makes or models and compare them side by side) or by selecting many attributes and comparing all retrieved cars side by side (e.g. Lycos Decision Guide). The last type of comparison presented a problem - all possible attributes were displayed and when too many cars were compared, the user had to scroll in two directions, which made efficient comparison difficult.

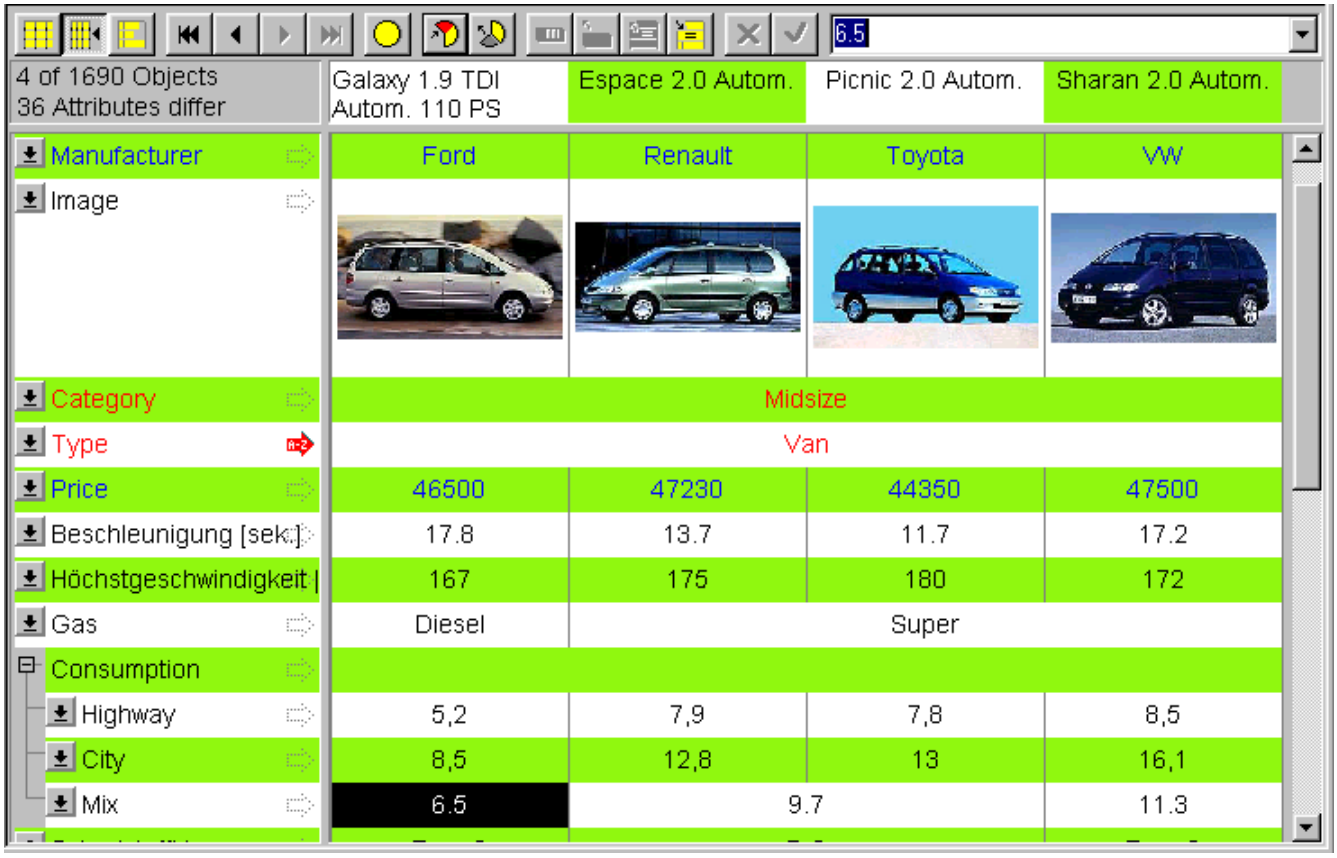

Figure 2: InfoZoom - Compressed Mode

3. INFOZOOM

InfoZoom's technique draws on Ahlberg and Schneiderman's [1] dynamic queries and Furnas' [6] fisheye techniques and allows the consumer to navigate the product space with a few mouse clicks. The consumer can get a flexible overview of an object–attribute table, which contains all products in the database (see Figures 1 and 2), and through selection of desired attributes and their values, restrict the table to the relevant set of data. This tool aims primarily at the Choice by Processing Attribute buying approach [15] but also allows direct access to products if the desired item is known.

InfoZoom displays database relations as tables with attributes as rows and objects as columns: in our case each column corresponds to a particular automobile. The attributes are hierarchically ordered like files in a directory.

The menu left of each attribute name shows the possible values and their frequency. Selecting a value from the menu restricts the table to the objects with this value. Clicking on the arrow outline right of an attribute sorts the table by this attribute.

In **Compressed Mode**, the column width is reduced until all the objects fit on the screen. In large tables the column width will be very small. Some techniques make the table readable in spite of this compression. The most important is that neighboring cells with identical values are combined into one larger cell. The width of each cell indicates the number of subsequent objects with this value. If a cell is to small to display a numeric value, a short horizontal line still indicates its relative height. Instead of selecting a value from the menu, an attribute can also be restricted by selecting and double-clicking a value or value-range directly in the table. In a short animation, the clicked cells grow while the others shrink. This looks like **zooming** into the table.

An **Overview mode** displays the entire database or subsections with all attributes sorted independently (Fig. 1). Missing values are also visible, so consumer can be aware of it and make a product evaluation on the basis of other attributes whereas in hierarchical interfaces products with missing values would be often excluded from the search results, treating "unknown" as "no match".

InfoZoom offers many ways of searching. The user can use free-string search, browse all products, select a set (or one product) according to desired criteria or combine both methods depending on search strategy and preferences. Users can select values from the display or attributes and value lists available for each attribute.

Because all possible characteristics are displayed the customer can select those characteristics that are the most important in the current task context rather than having to follow a preset sequence of choices. The customer can also select ranges for numeric values, depending on their need, which addresses the problem of fuzzy or arbitrary preset boundaries.

InfoZoom facilitates comparisons among products. Since products are displayed in a table customers can easily compare the value of various attributes, sort them or change the order of objects according to their needs.

Based on this analysis we predicted that users would perform well with InfoZoom as compared to standard hirarchical interfaces.

4. USABILITY STUDY

The feature comparison of InfoZoom and hierarchical interfaces suggested that InfoZoom should be a more efficient tool that facilitates searches for products. We conducted a comparative usability study to find out if potential EPC's users would confirm our analysis. In order to eliminate confounding factors such as network load, changes to content or site structures, and additional features such as advertising we decided to perform a controlled experiment with two handcrafted offline versions that resembled their online counterparts.

4.1 Materials

The database used for this study (AutoDB) contained 1690 cars described by over 20 attributes. The content was identical for both interfaces.

The hierarchical interface (HI) (Figs. 3-5) was implemented with forms in MS Access in a way to resemble the "Look & Feel" of typical online EPC catalog features. Because the site complexity and usability of the evaluated websites varied greatly, we tried to base our design on the most usable solutions that in our opinion performed well in various types of tasks. Our goal was to design a typical but advanced system with good usability for its class. The InfoZoom (IZ) interface was used in its unaltered format (see above, Figs. 1-2).

4.2 Participants

Test participants were recruited through flyers and word of mouth from universities, schools, companies, and neighborhoods in the Bonn, Frankfurt/Main, and Mannheim areas of Germany. 26 participants took part in the study, randomly assigned to either InfoZoom or the hierarchical interface. 15 participants were male and 11 users were female. Thirteen subjects were between 21 and 30 years old, the second biggest group (9 people) was comprised of 31 –40 year olds. One teenager and one older person (over 51) also took part in the study. A precondition for participation was (self-reported) Internet experience. Among the participants, sixteen (16) were experienced computer users (above 10h/week), six (6) worked on computers 3-9h/week, and four (4) could be classified as infrequent users with less than 2h of computer use per week. The distribution of experienced, intermediate and infrequent users turned out to be equal for both conditions. Seven (7) participants reported that they had purchased products (books, software, hardware, electronics) over the Internet, another six (6) users had searched the Web for information on products they intended to buy. Sessions took about an hour for which participants were rewarded with either a fashion watch or a prepaid vanity phone card, both valued under U.S.$15.

4.3 Tasks

The selection of tasks for the study was based on an typology of information seeking tasks, tasks supported by the sites we evaluated, and by informally interviewing people who had just bought a car or were about to buy a car to find out what kind of information they were looking for in their searches.

Tasks (T) were designed according to the following general characterization in increasing difficulty (ranking of degree of difficulty may, however, be subjective as some tasks were easier to perform on one interface than on the other and vice versa and much depended on the actual participant). We did pay attention that tasks were designed in a way so that car experts could not

guess the answers and a person who did not have much knowledge about cars could complete them. The tasks used were as follows (Questions have been translated here from the German versions used in the study):

- Find Attribute value for given Object
T1: *Does the Jaguar XJR have side airbags?*

- Find Database SubSet Attribute Value
T2*: How many different models does the Peugeot class 106 have?*
T4: *How many cars have the following characteristics: convertible, 100-150 HP, front drive, and cost 40.000 DM or less?*

- For given Attribute(s) value(s) Find Object(s)
T3: *Which middle class car is the cheapest one in the database?*
T6: *Which Funcar has the quickest acceleration?*

- Compare Attribute Values for 2 Objects
T5: *Which car has better fuel consumption: the Audi A6 1.8 or the Volvo S70 2.0?*

- Find Database Attribute Value
T7: *How many cars are in this catalog?*

- Compare Attribute Values for Subset
T8*: What is the price range (from the cheapest to the most expensive model) for the Alfa Romeo 145 series?*
T9: *Is the acceleration of the Mercedes-Benz S430 Automatic slow, normal, or quick for a ca rof its class?*

4.4 Procedure

Sessions were conducted by the first author at different places convienient for the participants (home, office, etc.) with the help of a laptop computer. At the beginning of each session participants were briefly introduced to the purpose of the study and the upcoming tasks. Participants filled out a short questionnaire about their demographics, computer experience, and e-commerce experience. Thereafter, participants in the InfoZoom condition were briefly informed about the interface followed by short (about 5 min.) demonstration of basic navigation features (selecting and marking values, zooming in etc.). The instructions were demonstrated on a different database than the one used for the study. Instructions were delivered that way so they did not suggest strategies for the tasks to be performed.

Because one of the preconditions to take part in the study was previous experience with WWW the basic interaction with the hierarchical interface (opening pull-down menus, selecting check-boxes etc.) was not practiced. No subject asked for clarification.

Next, participants were introduced to the nine tasks. After each task, they answered three questions, one each about their satisfaction, their perception of the time needed to accomplish the task, and their perceived efficiency by marking a 7-point Likert scale. Time to complete all tasks varied among participants from 11 to 45 minutes. After participants had concluded the last task they answered a usability questionnaire summarizing their experience. We used a slightly modified version of the CSUQ questionnaire [8]. The questionnaire was translated into German.

4.5 RESULTS

Results were analyzed for each category of data: accuracy, completion time, navigation, and satisfaction. Independent Samples t-tests were used for data analysis.

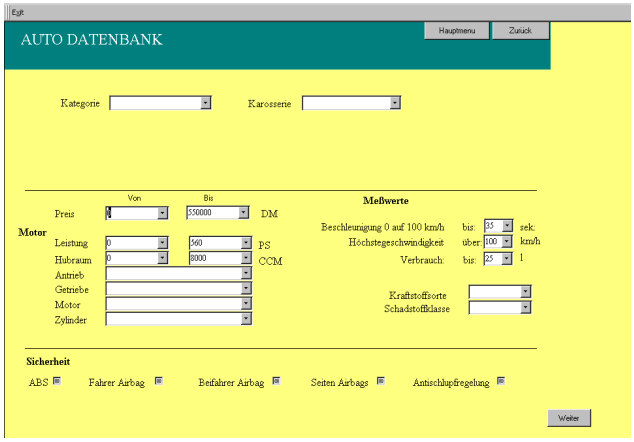

Figure 3: Hierarchical Online Catalog: Search Screen

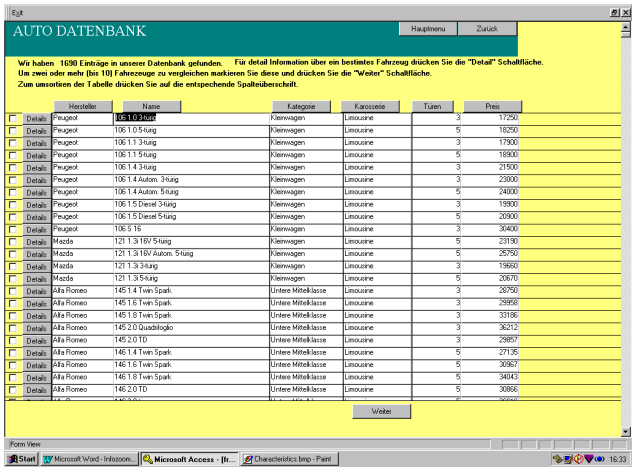

Figure 4: Hierarchical Online Catalog: Result List

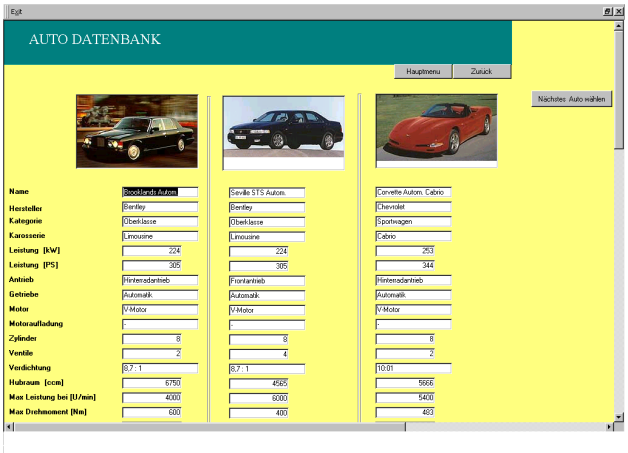

Figure 5: Hierarchical Online Catalog: Comparison

4.5.1 Completion rate and accuracy

Participants who evaluated InfoZoom finished all tasks while those working with the hierarchical interface did not complete the tasks in 9 cases (Table 2) which constituted a significant difference, χ^2=9.36, df=1, p<0.05. Participants encountered the biggest problems with task 7 (not completed in 4 cases) and task 9 (not completed in 2 cases); Non-completed trials were removed from time analysis. The number of incorrect answers (12) was equal for both interfaces, but the distribution over task types showed a different pattern. The data suggest that task T4, where participants were looking for the number of cars that fulfill four different criteria, was difficult for both groups. Task 7 – finding the total number of cars in the database - was difficult for the group evaluating hierarchical interface.

The most common mistake made by participants was assuming that data are sorted according to the variable they are actually looking for. This problem was common for both interfaces. Participants who evaluated the hierarchical interface had problems with ranges when both - lower and higher borders- had to be set and with the pull-down menus. In three cases they failed to scroll the menu to see additional choices. Those who worked with InfoZoom counted the pictures of objects (the same picture was sometimes used for different variants of the same car) instead of object IDs. One of the participants mistook the Overview display that has attribute values sorted independently (which means that columns are not meaningful) for a table.

Table 2: Number of incorrect answers and incompleted tasks for both interfaces.

Task	1	2	3	4	5	6	7	8	9	Sum
IZ-incorrect	0	1	0	6	0	1	0	4	0	**12**
HI-incorrect	1	3	0	4	1	2	1	0	0	**12**
IZ incomplete	0	0	0	0	0	0	0	0	0	**0**
HI incomplete	0	0	1	1	0	1	4	0	2	**9**
Sum	**1**	**4**	**1**	**11**	**1**	**4**	**5**	**4**	**2**	**33**

4.5.2 Completion Time

Time taken to complete each task was recorded for each interface. In six tasks the completion time was lower for InfoZoom interface. Total completion time for InfoZoom (AVG_{IZ}=19.4 Minutes, SD=6.9) was significantly faster compared to the Hierarchical Interface (AVG_{HI}=27.7 Minutes, SD=7.5) , t=2.921 df=24, p < .01.

In addition to recording actual time, time perception was reported at two different occasions: once after each task and later after the whole session. Comparing the data for actual time and time perception for each task (Fig. 6 and 7.), it can be seen that the values correlate with each other – the longer the actual time needed was for users to perform the task, the lower was their satisfaction with the speed of the performance, r=-0.456, df=224, p<0.01.

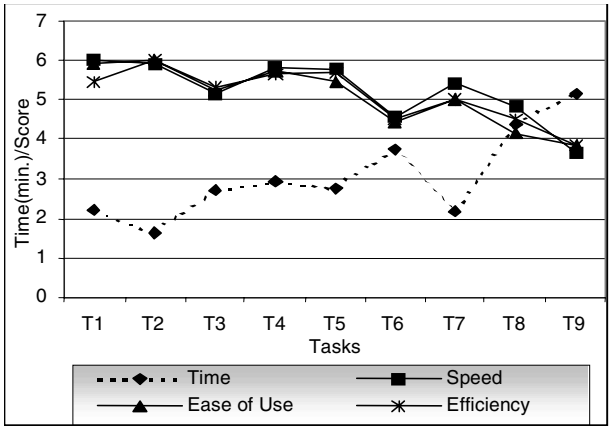

Figure 6: Hierarchical Interface: Time, speed, ase of use and efficiency

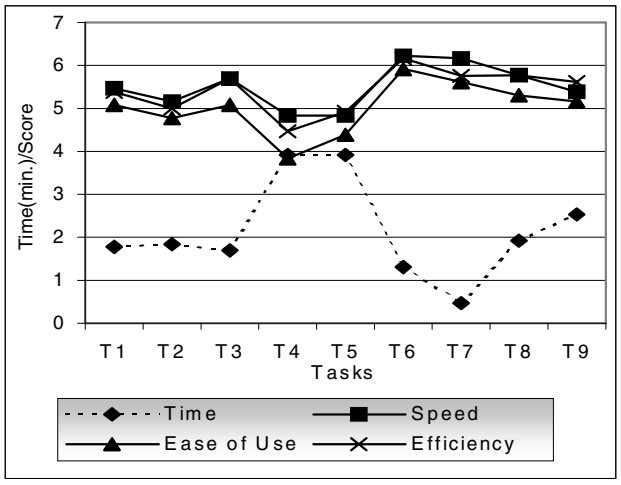

Figure 7: InfoZoom: Time , speed, ease of use and efficiency

4.5.3 Usability

Two parameters of usability were recorded after every task: ease of use and efficiency (Fig. 6 and 7). Overall, both parameters were valued higher for InfoZoom, users rating on average InfoZoom's ease of use 5.5 (out of 7) as compared to an average rating of 4.9 (out of 7) and 5.0 versus 4.8 for efficiency respectively, but these differences were overall not significant, t=1.76 (df=24), n.s. t-tests performed for single tasks revealed significant better ease-of-use ratings for InfoZoom on tasks T6 (t=3.074 (df=23), p< .01) and T9: (t=3.031 (df=22), p<.01). Similar results were obtained for efficiency.

Participants also judged other parameters of usability such as learnability, error recovery, interface organization, ease of searching, and interface features and the end of the trials. InfoZoom scored significantly higher (t=-2.712 df=21.991 p<.05) on "interface features" (number of features participants considered essential).

4.5.4 Satisfaction

After the session participants were asked 12 questions (on a 7-point Likert scale) about the interaction with the interface. Overall satisfaction with both systems was high. The Hierarchical interface was regarded as slightly more pleasant/comfortable than InfoZoom (AVG_{HI}=5.23 vs. AVG_{IZ}=5.15); users complained that they felt overwhelmed at first seeing the whole database. On all other questions users rated InfoZoom on average higher, significantly so on "fun" (AVG_{HI}=5.0 vs. AVG_{IZ}=6.3); (t=3.600 (df=23) p<.05) and finally overall satisfaction (AVG_{HI}=5.38 vs. AVG_{IZ}=6.38); t=-3.320,(df=24), p<.01.

4.6 OBSERVATIONS AND VERBAL PROTOCOLS

In addition to quantitative data we observed subjects throughout the tests. We noted users' strategies, critical incidents – breakdowns and things that went especially well - as well as comments made by users throughout the experiment.

4.6.1 Hierarchical Interface

Many observations made by participants about the hierarchical interface were consistent with our own observations during the o EPCs evaluation described above. Participants did not like the idea that they have to decide from the main menu in which part of the system (models or characteristics) they should look for desired information. Some seemed to be lost on pages that looked alike. Most of the participants did not read the instructions that were part of the interface. Instead, they would try to perform actions first and then read instructions in the case that their attempts did not produce desired results.

Participants seemed to be annoyed by the long pull-down menus. Only one person was using the typing of the first letter of the desired make to jump to the right part of the menu. Participants criticized lengthy result pages without the possibility of sorting attributes as well.

The page where users choose characteristics of the car (Fig. 3) caused the most problems for some participants. Some were not sure if default values should be deleted or left alone. Some were not sure if the values represented maximums and if the numbers represented in the pull down menus will retrieve all the cars from zero to the selected number or the cars from the smaller value on the list.

Some participants assumed that all the attributes visible on the page should have values chosen to make the system work properly. They attempted to guess values of other attributes not asked in the task. Many participants seemed to prefer to type values instead of scrolling pull-down menus, especially for numeric values. They complained that the numeric values in the pull-down menus were being set by the designer without the possibility for the users to select smaller subranges.

Much criticism was directed at the results pages. Participants complained that the choice of attributes visible for each car depended on the designer's choice (Fig. 4). Someone pointed out that it would be beneficial when the attributes that appear on the results page be those selected by the user on the previous page. Another person complained that there was only one level of sorting.

In the case when the participants selected criteria that did not match any car in the database, they received a message: "There is no car in the database that matches your criteria. Please choose another criteria and try again." When they saw the message twice, they usually stopped the search.

4.6.2 InfoZoom Interface

Participants working with InfoZoom frequently reported a feeling of initial information overload upon their first glance at the overview screen (Fig. 1), in part due to the large number of attribute available. They were also somewhat disturbed by the fact that for some attributes some values (character values) were visible, while those with fewer objects were represented by an empty selector or the first letters of the attribute value name. One user suggested to switch so as to give the user the possibility of choice between a proportional view (values grouped but not labeled) and a view where all the values would be visible, but still independently sorted.

InfoZoom allows applying many different search strategies to the search, depending on the user preferences. It was visible from the interaction that participants faced with a novel interface tended to choose one strategy and apply it to every task until they found out that this method was not appropriate for a particular task.

Participants who preferred selecting objects on the screen rather than using the Attributes or Values List complained about the difficulty to double-click on such small sectors like singular price value or car which contained only a couple of models.

Participants seemed to have also trouble with tasks that required comparison or gathering more than one different piece of information. The most common tactic used in such cases was to perform one part of the task and advance to another. But, at the same time, participants were aware that their method was not the most efficient one. After the session, they asked to have demonstrated the more efficient way. Some suggested the possibility of having two identical windows (or window split) to be able to compare different information pieces or the possibility to save the first selection to memory for future use.

Some participants seemed to be clearly aware about the program possibilities, even if they were not able to find it in the short time of the session or did not utilize them efficiently. For example, some participants asked questions about the string search possibilities, the possibility of comparing the objects by selecting many of them, the use of diagrams, the meaning of the icons on the toolbar, and so forth. Some discovered on their own features that were not demonstrated during the training session.

Since the study was carried out in 1999 many of the issues raised here have already been addressed in newer versions of the InfoZoom interface. A particular focus has been on simplifying the interface for the web context.

5. CONCLUSION

Both EPC's interfaces were rated highly, this shows that we created a reasonable baseline for comparison. InfoZoom was rated higher in overall performance. The observations made during evaluation of hierarchical sites were confirmed during the usability session – participants pointed out the same problems, and expressed a need for different solutions.

The results of our study suggest that InfoZoom is a promising solution for an interactive online catalog interface used in e-

commerce that addresses some of the shortfalls of the traditional interface approach.

The task type may be a major source of variance in customer response to interface style. InfoZoom performed considerably better for the last four tasks. Tasks 8 and 9, were the participants where comparing the object's attribute, with the subset when object and attribute was given, was significantly easier to perform with InfoZoom. This suggests that InfoZoom would support Choice by Attribute searching strategy better than hierarchical interface.

The increased performance for the last four tasks may also suggest easy learnability of the interface, especially since some participants expressed that the interface seemed overwhelming at the beginning, but after a few tasks was easy to work with. This interpretation can be supported by fact that T3 and T6 were constructed according the same schema and only the late one was performed significantly quicker and rated significantly higher and speed, efficiency. We could not find indications of a learning effect for the hierarchical interface. Future work will address the learnability issue in more detail.

The relatively small number of subjects and the between-group comparison design limited the power of our experiment, a counterbalanced within group study where a larger number of users learn and use both interfaces on multiple domains is desirable, but requires substantial investment in time –for subjects as well as experimenters.

This study has looked at catalog use in an experimental setting. Studies of e-commerce sites with different user incentives have shown that "pretend tasks" like the once here are only getting at parts of the picture, if you give people real money to carry out real shopping online you get different results (Jared Spool;UIE; personal communication). Similarily, features that were missing in this study such as advertising banners and special offers will affect and change the shopping (and buying behavior).

Nevertheless, since we focuses here on the product selection subtasks (not for example on shopping basket features) we think that lessons learned from this study can be transferred to online settings. More research on information seeking tasks and behavior will help us to better understand how people navigate online-shops and how we can design interfaces that are fun and help people find what they want.

6. ACKNOWLEDGMENTS

Infozoom is a product that originated at GMD – The German National Research Center for Information Technology and is presently further developed and marketed by humanIT Human Information Technologies GmbH. An evaluation version of Infozoom can be obtained from http://www.humanIT.com/.

The work reported here was carried out while the first author was a guest researcher at the GMD Institute for Applied Information Technology (GMD-FIT) and the second author was a member of GMD-FIT's research staff. We thank the staff of GMD-FIT for their support and suggestions throughout the study.

We also thank Andrew Dillon and Adam King and three anonymous reviewers for helpful comments and suggestions on earlier drafts of this paper.

7. REFERENCES

[1] Ahlberg, C., and Shneiderman, B. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing System: clebrating independence* (Boston MA, April 1994), 313-317.

[2] Belkin, N. and Croft, B.W. (1987). Retrieval Techniques. In Williams, M.E. (Ed.), Annual Review of Information Science and Technology (ARIST), Elsevier, 109-145.

[3] Bettman, J., and Pradeep Kakkar, Effects of Information Presentation Format on Consumer Information Acquisition Strategies. *Journal of Consumer Research,* 3 (March 1977), 233-40.

[4] Carroll, J and Rosson, M.B. (1987). The paradox of the active user. In J.M. Carroll (Ed.), Interfacing Thought: Cognitive Aspects of Human-Computer Interaction. Cambridge, MA: MIT Press.

[5] Dillon A., and Song M., An empirical comparison of the usability for novice and expert searchers of a textual and a graphic interface to an art-resource database. *Journal of Digital Information 1,1* (April 1997). http://jodi.ecs.soton.ac.uk/View/Index/v01/i01/

[6] Furnas, G. W. (1986). Generalized fisheye views. Proceedings of ACM CHI 86 Conference on Human Factors in Computing Systems.

[7] Harley Manning, John C. McCarthy, and Randy K. Souza Why Most Web Sites Fail, *Interactive Technology Series Volume 3, Number 7*, Cambridge, MA: Forrester Research, Inc., September 1998.

[8] Lewis, J. R., IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 1995, 7:1, 57-78

[9] Lohse, G.L.., and Spiller, P. Quantifying the Effect of User Interface Design Features on Cyberstore Traffic on Sales in *Conference proceedings on Human factors in computing systems* , (Los Angeles CA, 1998), 211 - 218.

[10] Maes,P. Guttman, R.H., and Moukas, A.G. Agents that Buy and Sell. Communications of the ACM, Vol.42, 1999 Nr. 3, pp. 81-91

[11] Newell, Allan and Herbert A. Simon. Human Problem Solving. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. 1972.

[12] Russo, J. Edward and Barbara A. Dosher. Strategies for Multiattribute Binary Choice, Journal of experimental Psychology: Learning, Memory, and Cognition, 9, 676-696.

[13] Schafer, J.B., Konstan J., and Riedl, J., Recommender Systems in E-Commerce. ; *Proceedings of the ACM Conference on Electronic Commerce*, (Denver CO, 1999), 158 - 166.

[14] Spenke, M; Beilken, Ch.; Berlage, Th., FOCUS: The Interactive Table for Product Comparison and Selection. in: *Proceedings of the UIST 96 Annual Symposium on User Interface Software and Technology* (Seattle, 6-8, 1996), ACM 1996, 41-50.

[15] Steiger P, Stolze M, Effective Product Selection in Electronic Catalogs, in: Extended Abstracts CHI97, ACM Press, 291-292

[16] Tilson, R., Dong, J., Martin, S., and Kieke, E., Factors and Principles Affecting the Usability of Four E-commerce Sites, in: *Proceedings of Fourth Conference on Human Factors and the Web*, AT&T LabsUSA, June 1998. http://www.research.att.com/conf/hfweb/proceedings/tilson/

8. APPENDIX

List of Websites Evaluated (Fall 1999)

1. www.autobytel.com
2. www.Auto.de
3. www.autoinfo.de
4. Automarkt
5. www.auto-shop.de
6. www.autosite.com
7. www.autotaal.nl
8. carpoint.msn.com
9. www.dadb.com
10. www.faircar.de
11. Lycos Decision Guide Cars
12. Mastercar.de
13. www.mobile.de
14. www.whatcar.co.uk