Usability of Information Visualization: Reading and Interaction Processes

Kasper Hornbæk

PhD Thesis
Department of Computing
Faculty of Science
University of Copenhagen
2001

Abstract

The usability of information visualizations is investigated in empirical studies of information retrieval, map navigation, and reading of electronic documents. Overall, subjects prefer using interfaces with overviews. However, analysis of the interaction processes show that subjects use mental and motor effort in switching to the overviews and that the overview occasionally distract the subjects. For some tasks, subjects using the overview are therefore slower. Zoomable user interfaces are faster than interfaces with overviews when subjects navigate on maps organized in multiple levels. We argue that reading of electronic documents is crucial for information access and use, and therefore aim at supporting that activity. An overview+detail interface for electronic documents improves the quality of essays that subjects write. Through visualizations of reading processes we describe how reading progresses and what parts of the documents subjects attend to. Subjects use an overview-oriented reading style to read electronic documents presented by a fisheye interface. Sections that the fisheye algorithm treats as unimportant are visible for a shorter time than in the other interfaces, although subjects feel uncomfortable in trusting the algorithm. In the studies described, different aspects of usability, such as efficiency and effectiveness, are not correlated. Consequently, we argue that studies of usability should measure a diversity of usability aspects. Finally, human thinking as described in introspective psychology is used to clarify designs of human-computer interaction and is suggested as a focus for further research in information visualization.

Dansk resume

Brugsvenligheden af informationsvisualiseringer er undersøgt i empiriske studier af informationssøgning, navigation på kort og læsning af dokumenter på elektronisk form. Samlet foretrækker forsøgspersonerne brugergrænseflader som præsenterer information ved et overblik kombineret med detaljer. Analyser af interaktionsprocesserne viser dog at skift til overblikket er mentalt og motorisk krævende og at overblikket til tider distraherer forsøgspersonerne. Ved nogle opgaver er forsøgspersoner som bruger overblikket derfor langsommere. Zoomende brugergrænseflader er hurtigere end brugergrænseflader med et overblik når forsøgspersonerne navigerer på kort organiseret i flere niveauer. Vi argumenterer for at læsning af elektroniske dokumenter har afgørende betydning for adgang til og brug af information, og søger derfor at støtte læsning. En overblik+detalje grænseflade forbedrer kvaliteten af essays som forsøgspersonerne skriver. Ved hjælp af visualiseringer af læseprocessen beskriver vi hvordan læsning skrider frem og hvilke dele af et dokument forsøgspersonerne koncentrerer sig om. Forsøgspersonerne anvender en overbliks-orienteret læsestrategi til at tilegne sig elektroniske dokumenter præsenteret med en fiske-øje grænseflade. De afsnit af dokumentet som fiske-øje algoritmen behandler som uvæsentlige er synlige i kortere tid end i de andre grænseflader, selvom forsøgspersonerne ikke føler sig trygge ved algoritmen. I de omtale studier er aspekter af brugsvenlighed, såsom effektivitet og produktivitet, ikke korreleret. Derfor argumenterer vi for at studier af brugsvenlighed skal måle en vifte af brugsvenlighedsaspekter. Endelig bruges menneskelig tænkning som beskrevet i introspektiv psykologi til at gøre klarere udvalgte design af menneske-datamaskine interaktion og foreslås som et fokus for yderligere forskning i informationsvisualisering.

Preface

This thesis is submitted to obtain the PhD degree at the Department of Computing, Faculty of Science, University of Copenhagen (DIKU). The work described in the thesis was carried out between November 1998 and November 2001.

The thesis concerns usability, reading, and interaction processes in information visualizations. It consists of two parts. The first part summarizes the contributions in the PhD. The second and main part of the thesis consists of six papers, referred to with the numbers 1 to 6. The papers are listed on page 5. The full papers are included from page 17 and on.

During my PhD work, I have enjoyed advice and inspiration from many scholars and scientists. First and foremost, I am grateful to my supervisor Erik Frøkjær for trying to teach me a consequent and engaged approach to doing research. In addition, Erik generously contributed his time and brainpower to discussing my every thought, ranging from the notion of knowledge to the maintenance of a used car. Ben Bederson, Catherine Plaisant, and Ben Shneiderman at the Human-Computer Interaction Lab at Maryland University have had a continuing impact on my understanding of good research. They and all other members of the Lab made my stay in the US very worthwhile, for which I am grateful. In addition to the friends and colleagues acknowledged in the papers, Peter Naur deserves special thanks for his support to emphasize clear thinking and writing, Morten Hertzum for educating collaboration, and Per Settergren Sørensen for patient and thorough advice on statistics. Finally, at DIKU good colleagues made the social and intellectual everyday-life pleasant—thanks.

Kasper Hornbæk,
Copenhagen, November 2001

Contents

List of Papers

Paper 1 (page 17-25):

K. Hornbæk and E. Frøkjær (1999), ‘Do Thematic Maps Improve Information Retrieval?’, in *Proceedings of INTERACT '99-Seventh IFIP TC13 Conference on Human-Computer Interaction*, Edinburgh, Scotland, 30th August- 3rd September 1999, p. 179-186.

Paper 2 (page 26-34):

E. Frøkjær, M. Hertzum & K. Hornbæk, (2000), ‘Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?’, *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2000)*, the Haag, the Netherlands, 1st-6th April 2000, p. 345-352.

Paper 3 (page 35-43):

K. Hornbæk & E. Frøkjær, (2001), ‘Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces’, *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2001)*, Seattle, WA, 31st March-5th April 2001, p. 293-300.

Paper 4 (page 44-74):

K. Hornbæk, B. Bederson & C. Plaisant (2001), ‘Navigation Patterns and Usability of Overview+detail and Zoomable User Interfaces for Maps’, 30 pages, a revised version to appear in *ACM Transactions on Computer-Human Interaction*. A previous version appeared as *HCIL Technical Report # 2001-11*, University of Maryland, ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/2001-11html/.

Paper 5 (page 75-101):

K. Hornbæk & E. Frøkjær (2001), 'Usability and Reading Patterns in Visualizations of Electronic Documents', 26 pages, under review.

Paper 6 (page 102-110):

E. Frøkjær & K. Hornbæk (2001), 'Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing', 8 pages, to appear in *Proceedings of HF2002/OzCHI*, Melbourne, Australia, 25th-28th November 2002.

1 Introduction

This thesis is about usability, reading, and interaction processes in visualizations for information access and use. In the following eight-page summary, I outline the background for the thesis, the aims of the thesis, and the contributions made in the thesis. Detailed results and discussions may be found in the papers beginning on page 17.

1.1 Background

User interfaces for information access and use help users find, manage, apply, and understand information. Such interfaces are well-known from digital libraries, the World Wide Web (WWW), geographical information systems, reference managers, and electronic books. They are discussed in the fields of Information Retrieval [Sparck Jones & Willett 1997], Information Visualization [Card et al. 1999], and Human-Computer Interaction [Baecker et al. 1995], among others.

User interfaces for information access and use are important for two main reasons. First, modern work requires large amounts of information. Managing and using this information have become increasingly difficult and in need of support. Estimates from a recent report [Lyman & Varian 2000] on the amount of information in the world will illuminate this development. In the year 2000, the information on the WWW has a size of approximately 4,200,000,000,000,000 bytes; between 610 and 1,110 billion emails were send; approximately 40,000 different issues of scholarly journals were published, many of them online; and around 7,500,000,000 original office documents were added to the already abundant paper and electronic archives. Most of this information must be accessed and used, making supportive interfaces indispensable.

Second, user interfaces for information access and use influence the outcome of the interaction and the interaction process itself. As one example, consider Superbook [Egan et al. 1989]—an interface with an expandable table of contents and string searching capabilities. In the final of three studies of the use of a statistics manual, subjects were faster, provided better answers to the questions posed, and were more satisfied with Superbook compared to a baseline interface. The interaction process in the final study also changed as a result of the new interface features. In addition to influencing the outcome and the interaction process, user interfaces may be more important than other parts of information access systems in determining the outcome of the interaction. Dumais [1996] argues that well-designed user interfaces consistently provide around 25% improvement in accuracy and speed, while other parts of information access systems, e.g. search algorithms, provide an average of 1-10% improvement. Thus, user interfaces are decisive in information access and use.

In information visualization, interactive visual representations on a computer are used to support human activities, especially information access and use. Information visualization has been pursued since the late 1980'es, originating at Xerox Palo Alto Research Center [Robertson et al. 1989; Card et al. 1991]. Among the systems proposed since are interfaces that support formulating queries [Young & Shneiderman 1993], give an overview of an entire collection of documents [Wise et al. 1995; Lin 1997], display search results [Nowell et al. 1996], support navigation within documents [Eick et al. 1992], assist relevance assessments of documents [Hearst 1995], and attempt to create an information work space [Card et al. 1996]. Through the nineties, a growing awareness has emerged of the need for empirical evaluation of information visualizations for information access and use, see Chen & Czerwinski [2000]. This awareness has lead to an increasing number of empirical studies of information visualizations [Beard & Walker 1990; North et al. 1995; Schaffer et al. 1996; Chen et al. 1998] and provides the background for my thesis.

1.2 Aims

This thesis aims at uncovering the usability of selected information visualizations for information access and use. I address some of the differences between visualizations called overviews, zoomable user interfaces, and fisheye interfaces. The intent is to establish a better understanding of the strength and weaknesses of these information visualization techniques and thus to assist designing more usable visualizations.

In addition to focusing on usability, the aim is to describe interaction processes in information visualizations. For example, I aim at describing how subjects read visualizations of documents. These descriptions are used to corroborate the usability measures obtained and to pose hypotheses about human thinking that might explain the interaction processes and the differences in usability between interfaces.

2 Contributions

The contributions of the thesis fall in four areas: (1) empirical data on benefits and drawbacks of three kinds of information visualizations—overviews, zoomable user interfaces, and fisheye interfaces, (2) guidelines on how to measure usability, (3) studies of interfaces for reading electronic documents and of how electronic documents are read, and (4) an analysis of user interfaces and central notions in Human-Computer Interaction in terms of human thinking.

2.1 Abstracts of papers

To provide an overview of the six papers comprising the thesis and to help readers understand the following discussion, I include below the abstracts for the papers.

Paper 1: *Do Thematic Maps Improve Information Retrieval?*

Thematic maps in the context of information retrieval are tools that graphically present documents and characterising terms. We investigated the usefulness of thematic maps in a laboratory experiment comparing a thematic map with a command language interface. Six subjects solved eight search tasks producing ten hours of logged and tape-recorded data. The experiment revealed no improvement in the quality of the documents retrieved when using a thematic map. A majority of the subjects considered the thematic map pleasant to use and thought that useful information was found on the map. However, searching took longer time using the thematic map compared with the boolean interface. Several subjects occasionally misinterpreted the structure and content of the map. The common expectation that thematic maps improve information retrieval lacks empirical underpinning and is in the present study only weakly confirmed.

Paper 2: *Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?*
Usability comprises the aspects effectiveness, efficiency, and satisfaction. The correlations between these aspects are not well understood for complex tasks. We present data from an experiment where 87 subjects solved 20 information retrieval tasks concerning programming problems. The correlation between efficiency, as indicated by task completion time, and effectiveness, as indicated by quality of solution, was negligible. Generally, the correlations among the usability aspects depend in a complex way on the application domain, the user's experience, and the use context. Going through three years of CHI Proceedings, we find that 11 out of 19 experimental studies involving complex tasks account for only one or two aspects of usability. When these studies make claims concerning overall usability, they rely on risky assumptions about correlations between usability aspects. Unless domain specific studies suggest otherwise, effectiveness, efficiency, and satisfaction should be considered independent aspect of usability and all be included in usability testing.

Paper 3: *Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces*
Reading of electronic documents is becoming increasingly important as more information is disseminated electronically. We present an experiment that compares the usability of a linear, a fisheye, and an overview+detail interface for electronic documents. Using these interfaces, 20 subjects wrote essays and answered questions about scientific documents. Essays written using the overview+detail interface received higher grades, while subjects using the fisheye interface read documents faster. However, subjects used more time to answer questions with the overview+detail interface. All but one subject preferred the overview+detail interface. The most common interface in practical use, the linear interface, is found to be inferior to the fisheye and overview+detail interfaces regarding most aspects of usability. We recommend using overview+detail interfaces for electronic documents, while fisheye interfaces mainly should be considered for time-critical tasks.

Paper 4: *Navigation Patterns and Usability of Overview+detail and Zoomable User Interfaces for Maps*
The literature on information visualization establishes the usability of overview+detail interfaces, but for zoomable user interfaces, results are mixed. We compare overview+detail and zoomable user interfaces to understand the navigation patterns and usability of these interfaces. The difference between these interfaces is the presence or absence of an overview of the information space. Thirty-two subjects solved navigation and browsing tasks on maps organized in one or multiple levels. Overall, users perform better with the multi-level map. We find no difference between interfaces in subjects' ability to solve tasks correctly. Eighty percent of the subjects prefer the overview+detail interface, stating that it supports navigation and helps keep track of their position on the map. However, subjects are faster with the zoomable user interface when using a multi-level map. The combination of the zoomable user interface and the multi-level map also improves subjects' recall of objects on the map. Switching between overview and detail windows was correlated with higher task completion time, suggesting that integration of overview and detail windows require mental and motor effort.

Paper 5: *Reading Patterns and Usability in Visualizations of Electronic Documents*
We present an exploration of reading patterns and usability in visualizations of electronic documents. Twenty subjects wrote essays and answered questions about scientific documents using an overview+detail, a fisheye, and a linear interface. We study reading patterns by progression maps that visualize the progression of subjects' reading activity; and visibility maps that show for how long different parts of the document are visible. The reading patterns help explain differences in usability between the interfaces and show how interfaces affect the way subjects read. With the overview+detail interface, subjects get higher grades for their essays. All but one of the subjects prefer this interface. With the fisheye interface, subjects use more time on gaining an overview of the document and less time on reading the details. Thus they read the documents faster, but display lower incidental learning. We also show how subjects only briefly have visible the parts of the document that are not initially readable in the fisheye interface. This happens even though subjects express a lack of trust in the algorithm underlying the fisheye interface. When answering questions, the overview is used for jumping directly to answers in the document and to already-visited parts of the document. However, subjects are slower at answering questions with the overview+detail interface. From the visualizations of the reading activity, we find that subjects using the overview+detail interface often explore the document further even when a satisfactory answer to the given question has already been read. Thus overviews occasionally grab subjects' attention and possibly distract them.

Paper 6: *Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing*
Understanding human thinking is crucial in the design and evaluation of human-computer interaction. Inspired by introspective psychology, we present five metaphors of human thinking. The aim of the metaphors is to help designers to consider important traits of human thinking when designing. The metaphors capture aspects of human thinking virtually absent in recent years of the CHI Conference Proceedings. As an example of the utility of the metaphors, we show how a selection of good and poor user interfaces can be appreciated in terms of the metaphors. The metaphors are also used to reinterpret central notions in human-computer interaction, such as consistency and information scent, in terms of human thinking. Further, we suggest the metaphors be used for evaluating interfaces.

2.2 Benefits and drawbacks of information visualizations

2.2.1 Overviews

The term overviews denotes two kinds of interfaces. One kind is usually called overview+detail interface and shows an overview of the entire information space together with a detailed view of the contents [Plaisant et al. 1995]. The other kind shows an overview of an entire document collection, for example in the form of a thematic map that shows the documents in the collection and words characterizing the main themes in the documents, e.g. Chen et al. [1998].

In our papers 1, 3, and 4[1], the most prominent benefit of overviews is that they increase satisfaction. In two studies, we found interfaces with overviews scoring higher on satisfaction questionnaires compared to alternative interfaces. In addition, subjects' preferences were consistently in favour of overviews, with 60%, 80%, and 95% of the subjects preferring the overviews. Subjects explained the satisfaction and preference data by saying that they (a) liked the overview of the structure of the information, (b) liked to use the overview for navigation, and (c) found the overview pleasant. These findings support data from experiments on overview+detail interfaces [North & Shneiderman 2000] and the literature on design [Greene et al. 1997; Shneiderman 1998].

In the study reported in paper 3, subjects wrote essays about scientific documents after reading with an overview, a fisheye and a linear interface. The overview improved grades with a medium effect-size, according to Cohen [1992] (i.e. by one half on a grading scale from zero to four). This finding shows that overviews may improve interaction qualitatively, possibly because of support for navigation and for memorizing headings and document structure.

In another study (paper 1), subjects used a thematic map of an entire document collection to solve information retrieval tasks. Subjects were inspired to use terms seen on the map in subsequent queries. Terms seen on the map were used in queries as often as terms seen in the full-text of documents.

We find a number of problems with overviews. Specifically for thematic maps (paper 1), subjects occasionally misinterpreted the structure of the map and they had difficulty in interpreting relationships between documents and terms on the map.

A surprising finding is that for some tasks and some information spaces, overviews lead to higher task completion times than interfaces without overviews. In the study of thematic maps described in paper 1, searching the map was 31% slower than using a command language interface. Our study of electronic documents (paper 3) showed that the overview+detail interface leads to 20% longer task completion time compared to a baseline linear interface, when subjects used the interfaces for answering questions. As described in paper 4, tasks were solved 22% faster on a map organized in multiple levels with a zoomable user interface compared to a overview+detail interface.

In the papers, we give two explanations for the time differences observed. First, the higher task completion time might be the result of motor and mental effort in switching between the overview and detailed information about the information space. On thematic maps (paper 1), the interaction process contained more shifts between different modes of interaction, such as querying or browsing the map, than did a command language interface. In the zoomable user interface experiment described in paper 4, we found that subjects who actively navigated on the overview window had higher task completion times. While we believe these observations have not before been made for overviews, they are similar to research which shows that combinations of modes leads to higher task completion time compared to individual modes, see Hertzum & Frøkjær [1996] and Raskin [2000].

Second, overviews may attract subjects' attention by appearing as an easy way to navigate and by creating associations for what to do next. In the analysis of reading patterns in electronic documents in paper 5, we argue that the availability of an overview often lead subjects to explore the document further even when they have already located

[1] Numbers refer to the papers listed on page 5. The papers are included from page 17 and on.

a satisfactory answer to the question posed. On thematic maps (paper 1), we observed that subjects sometimes browsed the map aimlessly, apparently loosing track of their task in face of the attractiveness of the map and the easy navigation it affords. Both these overviews contain readable information. In contrast, the zoomable user interface with an overview, described in paper 4, contained no readable information. With that interface, some subjects were able to ignore the overview and achieve task completion times similar to the interface without an overview.

Note that high task completion times may be desirable. In some cases more full exploration of information spaces or documents are preferable. In these cases task completion time could be considered an indicator of engagement. However, our data suggests a trade-off for overviews between satisfaction and task completion time.

2.2.2 Zoomable user interfaces

Zoomable user interfaces show information objects organized in space and scale and let users interact directly with the information space, mainly through panning and zooming [Perlin & Fox 1993]. In their simplest form, zoomable user interfaces are detail-only interfaces that allow zooming and panning. Paper 4 showed that zoomable user interfaces in this form have some advantages over interfaces with overviews, as mentioned above. Subjects were faster with the zoomable user interface when using a map organized in multiple levels. The combination of the zoomable user interface and a multi-level map also improved subjects' recall of objects on the map. With increased interactivity and navigational cues in the detail window, the overview becomes less important for navigational purposes. Shifting to the overview takes time and apparently also hurt subjects' memory for map locations.

However, despite being faster zoomable user interfaces in their simple form lead to lower satisfaction compared to interfaces with overviews.

2.2.3 Fisheye interfaces

Fisheye interfaces show only the parts of an information space with importance above some threshold [Furnas 1986]. Importance is determined a priori, for example by the structure of the information space, and with reference to the users' current view of the information space.

In paper 3 and 5, we found a fisheye interface to be 16% faster than the alternative interfaces when subjects read to understand the contents of a document. Also, subjects using the fisheye interface employed an overview-oriented reading style, spending more time to initially orient themselves in the document and less time to linearly read through the document.

A problem with the fisheye interface is that around half of the subjects expressed dissatisfaction with having to depend on an algorithm for determining what parts of a document are important. Whereas subjects expressed a lack of trust in the algorithm, they nevertheless used 30% less time compared to the other interfaces in sections that the algorithm determined to be unimportant. This behaviour may reflect a kind of premature cognitive commitment for some subjects—the fisheye interface apparently changes their perception of the document, even though they do not trust the algorithm. Subjects who used the fisheye interface answered correctly fewer incidental-learning questions after having read the document compared to subjects using the other two interfaces.

2.3 Measuring usability

In paper 2 we argued that usability testing of systems intended for complex tasks should measure both efficiency, effectiveness, and satisfaction. We analyzed how 87 subjects solved information retrieval tasks about programming problems. The correlation between efficiency, measured as task completion time, and effectiveness, measured as the quality of the solution, was negligible for practical purposes. Thus, we cannot a priori assume a certain relation between usability aspects—for example that fast interfaces are also effective. To show the practical consequences of this finding, we selected 19 papers from the CHI conferences. Of these, 11 measured only one or two usability aspects. The claims made in these papers about overall usability are thus weakened by the choice of usability parameters and could be plain wrong. We also showed how measuring all three aspects of usability helped the authors of one study to explain their surprising results.

Our other studies corroborate this finding. In all the empirical studies in this thesis, we measured all aspects of usability, as well as indicators of interaction and navigation processes. In every case we find that interfaces have high usability as indicated by one usability aspect, but low usability as indicated by another usability aspect.

2.4 Reading electronic documents

In paper 3 and 5, we argue that reading forms a crucial part of information access. Electronic documents are increasingly available during the information access process, for example on the WWW and in digital libraries. Users therefore have the possibility of reading electronic documents while they search and in that way resolve their information problems. Also, a large portion of the information access process consists of reading. In one study (paper 1), for example, subjects spend on average one-third of the information access process skimming and reading the full-text of documents, using what they read in formulating queries and judging the relevance of documents. In an empirical study of interfaces for reading (paper 3), we found that the most commonly used interface for electronic documents was inferior in most aspects of usability compared to interfaces based on information visualization. In summary, support for reading should receive much more focus in the design of interfaces for information access and use.

In paper 5, we use visualizations to closely study how subjects' reading activity progress and which parts of the documents subjects direct their attention toward. These visualizations serve to uncover that overview+detail interfaces may lead to further explorations of the electronic documents and that fisheye interfaces change how long subjects look at different parts of the documents. The visualizations of reading activity also show different modes in how subjects read the documents: subjects using a fisheye interface spend more time initially orienting themselves and less time reading linearly through the document. We also show how subjects used the overview area to return to previously visited places in the document and demonstrated reading behaviours such as flip-throughs, in which subjects navigate quickly through the entire document.

2.5 Human thinking as a focus for further work

In paper 6 we argue that a better understanding of human thinking is crucial in design and evaluation of user interfaces. However, central aspects of human thinking—such as knowing and habits—are virtually absent from papers in recent CHI conferences. To support more focus on human thinking in design and evaluation, we suggested five metaphors based on the works of James [1890] and Peter Naur [1995; 2000]. The metaphors help appreciate a selection of good and poor user interfaces. We also find that the metaphors clarify central notions in Human-Computer Interaction in terms of human thinking and help designers appreciate good and poor user interfaces.

The metaphors and the focus on human thinking are highly relevant to information visualization. The metaphors allow the notion of information scent [Pirolli & Card 1999] to be sketched in terms of human association. In addition, overview+detail interfaces and focus+context interfaces may be understood in terms of the metaphors. We also believe that central questions in information access, such as the relation between queries and information needs, can be clarified. In the study of zoomable user interfaces (paper 4), lack of support for habit formation posed problems to several subjects. Currently, we can only present these initial applications of descriptions of human thinking to visualization and information access. Thus, the ideas put forward in paper 6 serve mainly to create a focus for further work.

3 Conclusion

The usability of information visualizations has been investigated. In addition, reading and interaction processes were shown to be of crucial importance in interfaces for information access. For designers, we have shown how overviews incur trade-offs in usability, how to measure usability robustly, how reading may be better supported, and ways of considering human thinking in the design of user interfaces. For researchers, further challenges will be to replicate and extend this work in long-term studies of real-life tasks. Also, more research is needed on solid measures of usability, on improving the visualization techniques discussed, and on understanding individual differences in interaction and reading. Finally, we need to better understand interaction processes in information visualizations, e.g. the distraction triggered by overviews. To me, it does not seem to ambitious to seek a new theory of information visualization that explain interaction processes and usability in terms of human thinking.

References

Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S. 1995. *Readings in Human-Computer Interaction.* San Fransisco CA: Morgan Kaufmann.

Beard, D. B. and Walker, J. Q. 1990. Navigational Techniques to Improve the Display of Large Two-Dimensional Spaces. *Behaviour and Information Technology, 9,* 6, 451-466.

Card, S. K., Mackinlay, J. D. & Shneiderman, B. 1999. *Readings in Information Visualization.* San Francisco CA: Morgan Kaufmann.

Card, S. K., Robertson, G. G., & Macinlay, J. D. 1991. The information visualizer, an information workspace. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI'91, New Orleans, LA, Apr. 27- May 2). New York NY: ACM Press, 181-188.

Card, S., Robertson, G. G., & York, W. 1996. The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. In *Conference proceedings on Human factors in computing systems* (CHI'96, Vancouver, Canada, Apr. 13-17). New York NY: ACM Press, 111-117.

Chen, C. and Czerwinski, M. P. 2000. Special Issue on Empirical Evaluation of Information Visualizations. *International Journal of Human-Computer Studies, 53,* 5.

Chen, H., Houston, L., Sewell, R., and Schatz, R. 1998. Internet browsing and searching: User evaluations of category map and concept space techniques. *Journal of the American Society of Information Systems, 49,* 7, 582-603.

Cohen, J. 1992. A power primer. *Psychological Bulletin, 112,* 1, 155-159.

Dumais, S. T. 1996. Interface Design for Information Retrieval. In *Workshop on Social Aspects of Digital Libraries*. Available at http://is.gseis.ucla.edu/research/dl/dumais.html.

Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., and Lochbaum, C. C. 1989. Formative Design-Evaluation of SuperBook. *ACM Transaction on Information Systems, 7,* 1, 30-57.

Eick, S. G., Steffen, J. L., and Sumner, E. E. 1992. Seesoft-A Tool for Visualizing Line Oriented Software Statistics. *IEEE Transactions on Software Engineering, 18,* 11, 957-968.

Furnas, G. W. 1986. Generalized fisheye views. In *Conference proceedings on Human factors in computing systems (*CHI '86, Boston MA, Apr.), New York NY: ACM Press, 16-23.

Greene, S., Marchionini, G., Plaisant, C., and Shneiderman, B. 1997. Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information-Seeking. *Journal of the American Society for Information Science, 51,* 3, 380-393.

Hearst, M. A. 1995. TileBars: visualization of term distribution information in full text information access. In *Conference proceedings on Human factors in computing systems* (CHI '95, Denver CO, May). New York NY: ACM Press, 59-66.

Hertzum, M. and Frøkjær, E. 1996. Browsing and querying in online documentation: a study of user interfaces and the interaction process. *ACM Transactions of Human-Computer Interaction, 3*, 2, 136-161.

James, W. 1890. *Principles of Psychology*. Henry Holt & Co.

Lin, X. 1997. Map displays for information retrieval. *Journal of the American Society of Information Systems, 98,* 1, 40-54.

Lyman, Peter and Varian, Hal 2000. How much information? Retrieved from http://www.sims.berkeley.edu/how-much-info on 17/7 2001.

Naur, P. 1995. *Knowing and the Mystique of Logic and Rules.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Naur, P. 2000. CHI and Human Thinking. In *Proceedings of The First Nordic Conference on Computer-Human Interaction* (NordiCHI 2000, Stockholm, Oct. 23-25). Available from http://www.naur.com.

North, C. and Shneiderman, B. 2000. Snap-Together Visualization: Evaluating Coordination Usage and Construction. *International Journal of Human Computer Studies, 53,* 715-739.

North, C., Shneiderman, B., & Plaisant, C. 1995. User controlled overviews of an image library: A case study of the Visible Human. In *Proceedings of the 1st ACM International Conference on Digital Libraries* (DL '96, Bethesda, MD, Mar. 20-23). New York NY: ACM Press, 74-82.

Nowell, L. T., France, R. K., Hix, D., Heath, L. S., & Fox, E. A. 1996. Visualizing search results: some alternatives to query-document similarity. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval* (SIGIR'96, Zurich, Switzerland, 18-22 Aug.). New York NY: ACM Press, 67-75.

Perlin, K. & Fox, D. 1993. Pad: An Alternative Approach to the Computer Interface. In *Proceedings of the 20th Annual ACM Conference on Computer Graphics* (SIGGRAPH '93, Anaheim, CA, Aug. 2-6). New York NY: ACM Press, 57-64.

Pirolli, P. and Card, S. 1999. Information Foraging. *Psychological Review, 106,* 4, 643-675.

Plaisant, C., Carr, D., and Shneiderman, B. 1995. Image browsers: Taxonomy, guidelines, and informal specifications. *IEEE Software, 12*, 2, 21-32.

Raskin, J. 2000. *The Humane Interface: New Directions for Designing Interactive Systems.* Reading MA: Addison-Wesley.

Robertson, G. G., Card, S., & Mackinlay, J. D. 1989. The cognitive coprocessor architecture for interactive user interfaces. In *ACM SIGGRAPH symposium on User interface software and technology* (UIST'89, Williamsburg, USA). New York NY: ACM Press, 10-18.

Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., and Roseman, M. 1996. Navigating Hierarchically Clustered Networks through Fisheye and Full-Zoom Methods. *ACM Transactions on Computer-Human Interaction, 3,* 2, 162-188.

Shneiderman, B. 1998. *Designing the User Interface.* Reading, MA: Addison-Wesley.

Sparck Jones, K. & Willett, P. 1997. *Readings in Information Retrieval.* San Francisco CA: Morgan Kaufmann.

Wise, J. A., Thomas, J. J., Penncok, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. 1995. Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents. In *Proceedings of IEEE Symposium on Information Visualization* (InforVis'95, Atlanta GA, 30-31 Oct.). IEEE Press, 51-58.

Young, D. and Shneiderman, B. 1993. A Graphical Filter/Flow Representation of Boolean Queries: A Prototype Implementation and Evaluation. *Journal of the American Society of Information Systems, 44,* 6, 327-339.

Paper 1—Do Thematic Maps Improve Information Retrieval?

K. Hornbæk and E. Frøkjær (1999), 'Do Thematic Maps Improve Information Retrieval?', in M.A. Sasse & C. Johnson (eds) *Proceedings of INTERACT '99-Seventh IFIP TC13 Conference on Human-Computer Interaction*, Edinburgh, Scotland, 30th August- 3rd September 1999, Amsterdam: IOS Press, p. 179-186.

Human–Computer Interaction — INTERACT'99
Angela Sasse and Chris Johnson (Editors)
Published by IOS Press, © IFIP TC.13, 1999

Do Thematic Maps Improve Information Retrieval?

Kasper Hornbæk & Erik Frøkjær

Department of Computing, Universitetsparken 1, DK-2100 Copenhagen Ø, Denmark.

{kash,erikf}@diku.dk

Abstract: Thematic maps in the context of information retrieval are tools that graphically present documents and characterising terms. We investigated the usefulness of thematic maps in a laboratory experiment comparing a thematic map with a command language interface. Six subjects solved eight search tasks producing ten hours of logged and tape-recorded data. The experiment revealed no improvement in the quality of the documents retrieved when using a thematic map. A majority of the subjects considered the thematic map pleasant to use and thought that useful information was found on the map. However, searching took longer time using the thematic map compared with the boolean interface. Several subjects occasionally misinterpreted the structure and content of the map. The common expectation that thematic maps improve information retrieval lacks empirical underpinning and is in the present study only weakly confirmed.

Keywords: user study, information visualization, information retrieval, interaction, thematic map.

1 Introduction

This paper describes an exploratory investigation comparing a visual information retrieval interface (VIRI) and a command language interface. The background for this study is an appreciation of the importance of user interfaces in information retrieval (IR), the growing interest in VIRIs, and the few empirical studies of such interfaces.

The user interface of an IR system is of crucial importance to the interaction between user and system, and to the information retrieved. Different interfaces to the same search engine may lead the user to dissimilar results, and alter the search process. Special support of information searchers beyond traditional command language interfaces may increase searchers' performance and satisfaction by supporting formulation of queries and browsing of documents (Shneiderman, 1997; Hearst, 1999), and influence the number of queries formulated and the search strategies and tactics employed (Hertzum & Frøkjær, 1996).

In the 1990s there has been a continually growing interest in visual user interfaces to IR systems. VIRIs graphically display queries, documents, or meta information. Such interfaces have been expected to support formulation of queries, to facilitate browsing of document collections, and to support assessment of the relevance of documents. The usefulness of VIRIs supposedly is rooted in the characteristics of the human visual system, in the popularity and efficiency of graphical user interfaces generally, and in the concentration of information displayed in a VIRI.

Even though more than 20 interfaces satisfying the above definition of VIRI have been described in the literature — cf. Shneiderman (1997), Hearst (1999), few studies provide any empirical consolidation of the claims on the usefulness of visual interfaces. We expect that empirical investigations of the usefulness of VIRIs will raise new research questions and point to promising ways of improving VIRIs.

The remainder of the paper is structured as follows. We first delineate the work done on VIRIs and the underlying hypotheses about the advantage of VIRIs. The next two sections present the method used in our empirical investigation and the results of this investigation. The results are discussed and a conclusion is drawn.

2 Thematic Map Interfaces

Hearst (1999), Shneiderman (1997), and Gershon et al. (1998) describe VIRIs and the underlying assumptions. This paper focuses on a subset of VIRIs called thematic map interfaces. The term thematic map designates an information retrieval interface that depicts themes in a document collection on a two or three dimensional map, showing documents and characterising terms in an analogy to a geographical map.

The literature proposes numerous hypothesis about the general advantages of VIRIs and the more specific benefits of thematic map interfaces. VIRIs are claimed to improve the quality of the search, to improve the search process, and to improve the subjective satisfaction with the information retrieval system. Especially, in virtue of the overview produced by a thematic map, the quality of open-ended or explorative information retrieval tasks is believed to be supported by VIRIs (Chalmers & Chitson, 1992). It is also commonly assumed that searchers will perform faster because they rely on their perceptual rather than their cognitive capabilities (Korfhage, 1991). With respect to the search process, it has been argued that VIRIs will support users in their initial orientation in a system and in their endeavour to express their information needs (Shneiderman, 1997). The graphical arrangement of documents and terms on a map is expected to support decision on whether or not a document is relevant. Thematic maps are also thought to inspire the user in finding documents that would otherwise have been unnoticed (Lin, 1997). VIRIs are also conjectured to increase subjective satisfaction. These claims will form the hypotheses of the empirical investigation described in Sections 3 and 4.

2.1 Previous Work

Thematic map interfaces were first introduced to IR by Xia Lin — cf. Lin (1997). Lin used an algorithm devised by Kohonen to construct a two dimensional representation of document collections. The technique was demonstrated using two collections, one indexed by 140 titles, one by 660 titles, keywords, and abstracts. Through an iterative procedure, the Kohonen algorithm organizes documents using similarities in the words occurring in the documents. Major themes in the collection are extracted and the documents grouped according to those themes. Lin's thematic map interface shows documents and terms in distinct areas, where each area is characterised by the term occurring most frequently in the documents in that area. Consequently, the map is thought to convey information about salient terms and the overall structure of the document collection. In a recent experiment, Chen et al. (1998) created a thematic map interface to 110,000 web pages. Furthermore, Chen and his colleagues added some interactivity to the Kohonen map by making it possible to click on an area of the map and get a new thematic map, showing themes only from the web pages in that area.

Another group of thematic map interfaces is based on multidimensional scaling (MDS), a family of statistical projection techniques that maps documents into low dimensional space (Chalmers & Chitson, 1992; Wise et al., 1995). In the BEAD system documents are distributed in a three dimensional space using physical modelling of documents (Chalmers & Chitson, 1992). So-called forces between documents are calculated using keywords in the documents thus grouping documents according to themes. Wise et al. (1995) also uses multidimensional scaling to create a topological map of themes in a large document collection (>20,000 documents). The map shows themes in the corpus, with related themes adjacent. The system described in Wise et al. (1995) also shows the strength of the different themes in the collection as the height of the topological structures representing themes.

2.2 Empirical Investigations

Few empirical investigations of VIRIs have been published and only a couple treat thematic maps. Lin investigated how 68 users solved simple search tasks on different kinds of paper-based thematic maps — cf. Lin (1997). With respect to search time Lin concluded that the Kohonen-map was as good as a humanly constructed map for locating titles and significantly better than a random arrangement of documents.

Chen et al. (1998) made a comparison of a thematic map with browsing the hierarchical structure of the Internet search site Yahoo. 31 subjects tried to locate a web-page which contained "something of interest to you" (Chen et al., 1998, p.587). First the subjects tried to retrieve an interesting page using either the thematic map or the hierarchical structure. Afterwards subjects were to repeat the search task using the other interface. Chen et al. found that subjects were able to browse a thematic map and locate relevant information. However, searching in the map after a page already found was inefficient. Chen et al. also found that subjects seemed to like the graphical aspects of the map and thought that browsing using a map was a convenient way of searching for information. However, some subjects had difficulties in understanding the map and the words on the map.

As a consequence of the meagre empirical understanding of thematic map interfaces, the hypotheses mentioned above are largely untested. In a recent review, Gershon et al. (1998) point out that we need to make information visualization systems that are easy to use. Further, Gershon et al. argue that we should design human- and usage-centred information visualizations. This challenge is being faced here, based on the assumption that empirical knowledge about the use of VIRIs is necessary for designing useful visual interfaces.

3 Experimental Method

In order to study differences in the interaction process between a non-graphical information retrieval interface and a VIRI, a command language and a thematic map interface were constructed. The command language interface allowed subjects to formulate queries using boolean logic, to scan the result of a query, and to inspect full-text. In the following this interface is called the boolean interface. The thematic map was constructed using multidimensional scaling. In addition to the thematic map the VIRI had exactly the same functionality as the boolean interface. Therefore, observed differences in searching behaviour and in the search results between the two interfaces are attributable to the presence or absence of the thematic map.

The experiment was conducted employing a within-group design with interface type as the independent variable. Six subjects participated in the experiment solving eight tasks each, four task with each interface. The order of tasks as well as the order in which the subjects experienced the two interfaces were alternated, minimising learning effects. The hypotheses for the experiment were taken from the literature, outlined in Section 2. Tasks were given to the subjects on separate sheets concisely describing the search task. Four of the experimental tasks explicitly described the documents that were to be found and what would count as a satisfactory answer, for example "Find the paper by Rudolf Darken on wayfinding in virtual worlds." The remaining four tasks were aimed at a broader group of documents and could be answered in more diverse ways, e.g. "Imagine that you are to give a talk on the use of computers in education. What is available on that topic?" The experiment was conducted in a dedicated lab with subjects who were master thesis students in computer science. All subjects had self-acclaimed knowledge about human-computer interaction, the subject area of the document collection used, and all had experience in using boolean logic in IR systems.

During the experiment, queries, inspection of full-text, and interaction with the map were logged. The subjects were encouraged to think aloud while searching. The think aloud utterances were recorded on tape. Before solving the search tasks, the subjects were interviewed concerning their personal and educational background and search experience. A short post-search interview about the satisfaction with and usefulness of the two interfaces was also conducted.

3.1 Boolean Interface

The boolean interface is shown in Figure 1. The user may formulate queries using search terms in combination with the boolean operators AND, OR, and NOT. The documents retrieved in response to a query are shown in an unranked list. The full-text of documents can be displayed by double clicking on the titles/authors of the retrieved documents. The full-text is automatically formatted using information about document structure; the appearance of the documents may therefore be different from the original article or conference paper. Full-text is presented within one second.

The experimental interfaces give access to 436 documents from conferences and journals on human-computer interaction. The documents were taken from HCILIB, an experimental IR system developed at University of Copenhagen (Perstrup et al., 1997). The documents accessible through the thematic map and boolean interface were indexed using full-text where non-content bearing words (stop-words) had been removed and the terms stemmed — see Salton & McGill (1983) for a description of these standard information retrieval techniques.

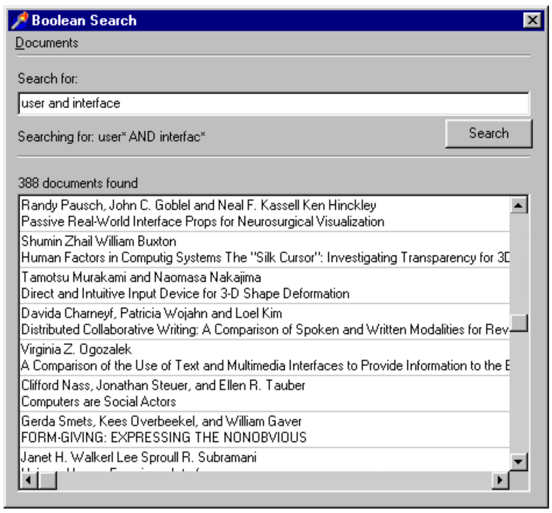


Figure 1: The boolean interface used in the experiment. In the edit box in the upper part of the screen search terms and boolean operators may be entered. Below the edit box are shown the stemmed terms used in the search. On the lower part of the screen are shown the author and title of the retrieved documents in an unranked list. If one clicks on one of the titles the full-text is shown.

3.2 Thematic Map Interface

The thematic map was constructed using multidimensional scaling (MDS) (Borg & Groenen, 1997). The MDS algorithm constructs a two dimensional arrangement of documents using a measure of similarity between documents. Document similarity was calculated from counts of words in the documents, using the cosine similarity measure (Salton & McGill, 1983). The MDS algorithm calculates the two dimensional arrangement through minimising the difference between the original inter-document similarity and the distances between documents in the two dimensional arrangement. The resulting thematic map is shown in Figure 2.

Terms describing themes were placed on the thematic map together with the documents. The terms on the map were selected by first calculating the discrimination value (Salton & McGill, 1983) for all terms and then placing the 20 terms with the largest discrimination value on the map. Intuitively, a term having a high discrimination value occurs frequently in some documents and rarely in others, for example 'evaluation' in this document collection. Such terms is here used for describing documents in which they frequently occur. The position on the map of the individual terms were found by calculating the 'mass' midpoint of the square of the number of occurrences of the term in all documents. The most frequent stem of a term was used as the actual text displayed on the screen.

The thematic map offers several ways of interacting with the VIRI that link the thematic map, the document list, and the query text. All documents on the map retrieved by a query are coloured yellow. Likewise, all terms on the map occurring in a query are coloured yellow. If the user selects one or more documents on the list of retrieved documents, the position of those documents are displayed on the map by a different mark than other documents. If the mouse is moved over a document on the map, the title and author of the document is shown in a pop-up box. It is also possible to right click with the mouse on a document on the map to see the full-text of documents.

If the user wants to enlarge a portion of the map it is possible to zoom. Zooming is smooth and is accomplished by holding down the left mouse button. It is always possible to zoom out to see the entire document collection. There is constantly about 20 terms on the visible portion of the map. When a user zooms on the map more terms with a decreasing discrimination value become visible. The thematic map described here employs a wider range of interaction techniques than maps described in the literature. A wider range of interaction techniques was suggested by Chen et al. (1998) as a way to improve thematic maps.

4 Results

In analysing the behaviour of the subjects the data logs were integrated with the verbal protocol. The statistical analysis was done using analysis of variance (ANOVA) and t-tests after removal of persistent differences between subjects and between different search tasks — cf. Hertzum & Frøkjær (1996). The quantitative analysis focused on confirming the qualitative results and describing search behaviour on the thematic map.

4.1 Documents Retrieved and Search Time

Table 1 shows the number of documents marked as relevant by the subjects upon using the two interfaces. There is no significant effect of interface type upon the number of documents retrieved ($F(1, 0.07), p > 0.79$). Nor is there any significant difference in the number of documents marked as relevant between the two interfaces ($F(1, 0.36), p > 0.55$). The relevance was judged by the first author. There is no significant difference between interfaces either, if relevance is approximated as documents marked relevant by more than one subject ($F(1, 0.32), p > 0.57$).

The time taken to complete the search tasks is shown in Table 2. Subjects use significantly longer time in the visual information retrieval interface compared with the boolean interface ($t = -2.975, p < 0.01$). There are also large individual differences in task completion time; averaged over the eight tasks the slowest subject took twice the time of the fastest. Within individual tasks solved using the same interface, task completion time differ by a factor of seven.

	Interface type	
	Boolean	Visual
Relevance	(N = 113)	(N = 108)
Relevant	66% (75)	77% (83)
Partial relevant	26% (29)	16% (17)
Non-relevant	8% (9)	7% (8)

Table 1: Documents marked as relevant by the subjects in the two interfaces. Relevance is expressed as one of three levels: relevant; partial relevant, for example documents about multimedia in response to a task on interfaces using sound; and non-relevant, that is documents containing no information relevant to the task.

Notable was the large proportion of time subjects used on scanning full-text. On the average one-third

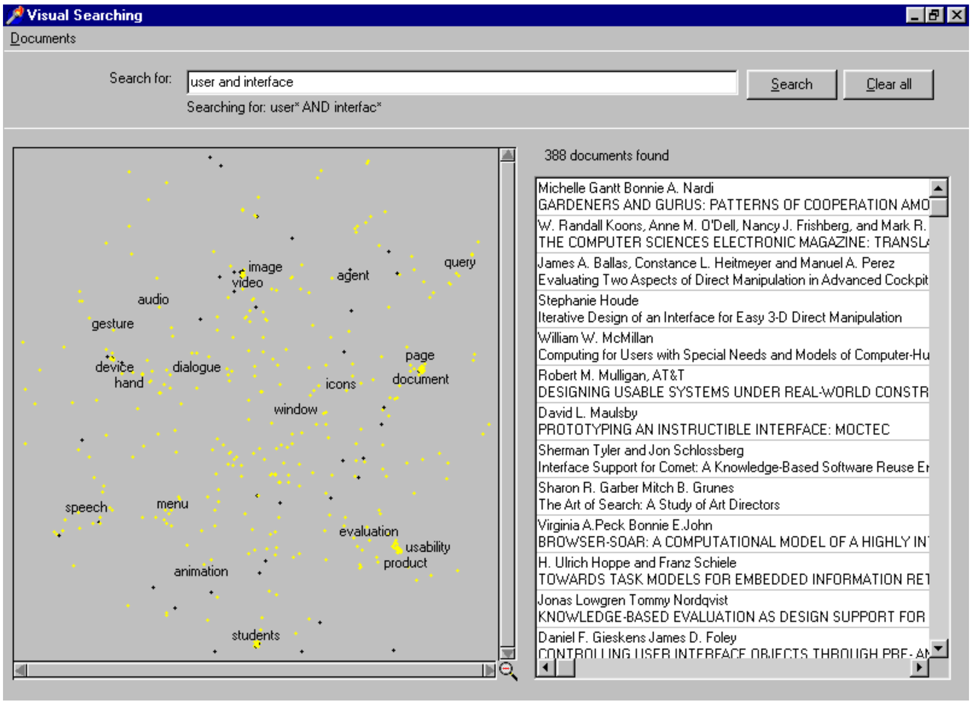


Figure 2: The thematic map interface used in the experiment. The input area in the upper part of the display as well as the list of document titles and authors to the right are identical to the boolean interface. On the map to the left documents are shown as dark and bright dots. The terms on the map are supposed to describe the contents of the documents around them. Documents retrieved by entering the query 'user and interface' are bright (yellow) on the map. The scrollbars next to the map allow navigation when there is zoomed on a region of the map.

of the time used for searching was spend on inspecting full-text, trying to judge relevance of the document or to locate useful search terms.

Time per task	Interface type Boolean	 Visual
Mean	10.8 (6.2)	14.2 (7.0)
Minimum	1	2
Maximum	22	25

Table 2: Time elapsed searching per task, in minutes. Searching in the visual interface is significantly slower than using the boolean interface. 24 tasks were done in each interface. Standard deviation is given in parenthesis.

4.2 Queries and Terms

The difference in the use of queries in the two interfaces is shown in Table 3. There is no statistically significant difference between the numbers of queries issued in the two interfaces ($F(1, 0.16), p > 0.6$), but there is a tendency towards issuing less complex queries using the thematic map ($t = 1.91, p < 0.07$). The average number of constituents of a query (counting terms and boolean operators), was 2.8 in the thematic map as against 3.4 in the boolean interface.

No. queries	Interface type Boolean (N = 105)	 Visual (N = 112)
Mean	4.4 (4.0)	4.7 (3.2)
Minimum	1	0
Maximum	17	11
Average number of constituents	3.4	2.8

Table 3: Queries in the boolean and visual interfaces. The table shows the average number of queries and query constituents, i.e. terms and operators, used in solving a task. One search task was solved exclusively using the thematic map, hence zero as the minimum number of queries using the map.

The think aloud protocol has been analysed as to where the inspiration to search terms came from. Inspiration to search terms is divided into four categories: 1) inspiration from the text describing the search task, 2) inspiration from association or ways invisible in the think aloud protocol, 3) inspiration from the titles and full-text of documents, and 4) inspiration from terms on the thematic map. Table 4 shows the distribution of term inspiration. The thematic map seems to inspire subjects as often as do the full-text or title of a document. Such inspiration would typically involve the user seeing a term on the map and then using that term in a query.

	Interface type	
Inspiration to terms from:	Boolean (N = 101)	Visual (N = 91)
Task description	56% (57)	51% (46)
Association	34% (34)	30% (27)
Title/Full-text	10% (10)	9% (8)
Thematic map	–	11% (10)

Table 4: Inspiration to search terms. The table shows the number of search terms unique to each solution of a task divided between different sources of inspiration.

4.3 Interacting with the Thematic Map

The subjects' use of the thematic map, scanning of titles, and inspection of full-text varied between the interfaces. This search behaviour may be described in terms of interaction shifts. An interaction shift is a change from one interaction mode (e.g. formulating queries) to another (e.g. scanning titles), as it can be detected from the log and the verbal protocol. Significantly more interaction shifts happen when using the VIRI than when using the boolean interface ($t = -2.957, p < 0.01$). This is partly because 14 out of 24 tasks in the boolean interface was solved by issuing one or more queries and then inspecting titles and full-text. That way of solving a task involves only one interaction shift, while the average number of shifts in the VIRI were six. The analysis of interaction shifts also show the different use of queries in the two interfaces. With the boolean interface the queries occur in sequences of average length 2.8, while with the visual interface queries are interwoven with browsing on the map and inspection of titles and full-text (the average number of queries without interaction shifts is 1.5).

Browsing on the map was preferred to scanning list of titles/authors. In two out of three searches the interaction with the visual interface started with the formulation of a query; the remaining tasks were begun by browsing the map. In the cases where the interaction with the visual map began with a query the first interaction shift often lead to the map; in 14 out of 24 search tasks the map is thus preferred to scanning the list of titles and authors.

In the 24 tasks solved with the visual interface, subjects directly interacted with the map in 16 tasks. It is difficult to quantify and evaluate search behaviour on the thematic map. There are, however, four prominent features of the use of the thematic map. First, when browsing the map subjects tend to focus on specific words or areas. In 14 out of 16 tasks solved with the aid of the map, the subjects focused on a word that was thought pertinent to the search task. During a task concerning sound in user interfaces, one subject said:

> "User interfaces using sound … then there was, what was it I found … it was called 'audio' and 'speech' on the map, because there are such words there [on the map], I think I'll zoom in and look if there is something."

There were also subjects who assumed that documents relevant to the task should be found in one specific area of the map. Focus on a particular area was observed especially when there was a large proportion of hits in one area of the map. Referring to a small area containing a lot of retrieved documents one subject said, "I'll just try to look at that cloud over here [on the map], to see why they are placed over here".

Second, there were several examples where subjects used the position of a document to judge its relevance, and where subjects used the position of a document on the thematic map to find other relevant documents. Pondering the relevance of a paper called "Relief from the audio interface blues" one subject said: "Well, I would say it [the document] is relevant, because it is next to the other [documents on the map judged relevant]".

Third, in a number of cases the interpretation of the map and of the relation between documents found adjacent on the map was haphazard. One example of such interpretation occurred when one subject focused on the rim of an area containing a lot of retrieved documents. The subject did so searching for documents on practical applications of GOMS. Since GOMS is a theoretical model the subject thought that relevant documents would be on the border of that area.

Forth, several of the subjects lost track of their task and browsed the map in a aimless way. This phenomena was primarily seen in searches for well-specified documents, where some subjects — when they couldn't find the document satisfying the task description — looked several times at the same areas and documents.

4.4 Subjective Satisfaction

Four out of six subjects expressed preference for the thematic map interface over the boolean interface;

they found the graphics pleasing, liked the overview gained from browsing the map, and found inspiration to formulating queries from the terms on the map. The following quotes describe this: "I preferred the graphical, it was more fun in some way. That's probably the best part about it [the thematic map]", and:

> "On the one hand you've got the words on the map and you can see how many documents you've retrieved, so it was faster to get an overview of your search: did you retrieve few or many, how are they placed in relation to each other [on the map], are they close to each other or more scattered."

One subject preferred the boolean system only because the window showing titles and authors was re-sizeable, and one subject found the thematic structure too difficult to understand.

Half of the subjects in the post-search interview expressed difficulties in understanding the map. The relation between terms and documents on the map was thought to be unclear as was the thematic structure of the map. One subject commented:

> "I'm wondering about the categories shown, they are a bit … some of them are main themes in computer science like 'evaluation' and 'usability' that one can relate to but something like 'hand' … that can mean anything."

Also several subjects expressed surprise when they inspected documents adjacent on the map and could not tell what the documents had in common. One subject remarked in the post-search interview: "One hopes that when they [the documents] are close they are about the same." Asked if documents adjacent did share a common theme the subject continued: "Perhaps half of the times".

5 Discussion

The hypotheses about VIRIs, outlined in Section 2, are only weakly confirmed in this experiment. Thematic maps used for IR did not improve the quality or number of documents retrieved, nor was searching faster.

However, thematic maps improve certain aspects of IR: users find searching on the map pleasant, prefer browsing to scanning list of titles, and get inspired to search terms from the map. Why, then, are the results of the retrieval process not improved? One explanation might be that users lose focus on the search task, given the number of interaction shifts between the thematic map and issuing queries, and the aimless browsing on the map observed in some search tasks. Both these distractions may also result in a time overhead compared with the boolean interface. Similar problems have been observed in other empirical studies. Hertzum & Frøkjær (1996) found that search time was negatively influenced by the availability of several interaction modes. Chen et al. (1998) reports that some users browsed the thematic map in an aimless way; aimless browsing is also reported in the literature on hypertext usage. These problems might be inherent in the graphical, non-sequential presentation of documents and in the combination of browsing and query use.

One way of improving IR with thematic maps might be to increase the understandability of thematic maps. This study documents that users experience problems with understanding the terms on the map and the relations between documents. The difficulty with understanding terms could be addressed by adding more context to the terms presented on the thematic map, e.g. by using phrases, sentences, or groups of words. The understandability of relations between documents might be improved by introducing explicit connections between documents presented, as in networks showing documents and terms (Fowler et al., 1991). Several subjects wanted the possibility of getting a part of the map presented as a list of document titles/authors. They argued that the manageable, linear structure of a list in certain situations was preferable to the associative structure of the map. Whether or not such changes to thematic maps will improve IR remains to be empirically investigated.

The present experiment supports the integration of browsing using a VIRI and searching using queries. Querying was used with the same frequency with the thematic map and the boolean interface. Some search tasks were successfully solved only using the querying function of the map. Other work comparing searching and browsing has also reached this conclusion (Hertzum & Frøkjær, 1996). It is much too simplistic to assume that IR can be improved using a browse-only thematic map, as in Lin (1997) and Chen et al. (1998).

The generality of the above conclusions may be questioned because of the relatively small number of subjects and the unrealistic experimental situation. Thus, further experiments should include more subjects, investigate support for complex tasks developing over time, and address the use of thematic maps by subjects experienced with such interfaces.

6 Conclusion

Contrary to the expectations raised in the literature, this study did not find any quantitative improvements of information retrieval using a thematic map. However, subjects prefer the thematic map compared with the boolean interface. The thematic map is also extensively used in the information retrieval process, for instance in finding useful search terms. A problem with thematic maps is the distraction caused by unfocused browsing and by shifts between different interaction modes. The thematic map was also misinterpreted with respect to relations between documents, and the significance of terms displayed on the map were not directly understandable.

In brief summary, this study and the few other studies of thematic maps have shown that far more work is needed to really improve information retrieval by thematic maps.

Acknowledgements

We would like to thank the HCILIB-group for granting us access to the documents used in this study. We also acknowledge comments on a earlier version of this paper by Peter Naur, Kristian Bang Pilgaard, and Ketil Perstrup.

References

Borg, I. & Groenen, P. (1997), *Multidimensional Scaling*, Springer-Verlag.

Chalmers, M. & Chitson, P. (1992), BEAD: Explorations in Information Visualization, *in* N. Belkin, P. Ingwersen & A. M. Pejtersen (eds.), *SIGIR'92 Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, pp.330–7.

Chen, H., Houston, A. L., Sewell, R. R. & Schatz, B. R. (1998), "Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques", *Journal of the American Society for Information Science* **49**(7), 582–603.

Fowler, R. H., Fowler, W. A. & A.Wilson, B. (1991), Integrating Query, Thesaurus and Documents through a Common Visual Representation, *in* A. Bookstein, Y. Chiaramella, G. Salton & V. Raghavan (eds.), *SIGIR'91, Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, pp.142–51.

Gershon, N., Eick, S. & Card, S. (1998), "Information Visualization", *Interactions* **5**(2), 9–15.

Hearst, M. A. (1999), User Interfaces and Visualization, *in* R. Baeze-Yates & B. Ribeiro-Neto (eds.), *Modern Information Retrieval*, Addison–Wesley.

Hertzum, M. & Frøkjær, E. (1996), "Browsing and Querying in Online Documentation: A Study of User Interfaces and the Interaction Process", *ACM Transactions on Computer–Human Interaction* **3**(2), 136–61.

Korfhage, R. R. (1991), To See, or Not to See — Is *That* the Query?, *in* A. Bookstein, Y. Chiaramella, G. Salton & V. Raghavan (eds.), *SIGIR'91, Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM Press, pp.134–41.

Lin, X. (1997), "Map Displays for Information Retrieval", *Journal of the American Society for Information Science* **48**(1), 40–54.

Perstrup, K., Frøkjær, E., Konstantinovitz, M., Konstantinivitz, T., Sørensen, F. & Varming, J. (1997), A World Wide Web-based HCI-library designed for interaction studies, *in Proceedings of the 3rd ERCIM Workshop on User Interfaces for All*, INRIA Lorraine, pp.137–42.

Salton, G. & McGill, M. (1983), *Introduction to Modern Information Retrieval*, McGraw-Hill.

Shneiderman, B. (1997), *Designing the User Interface: Strategies for Effective Human–Computer Interaction*, third edition, Addison–Wesley.

Wise, J., Thomas, J., Pennock, K., Lantrip, D., Porttier, M., Schur, A. & Crow, V. (1995), Visualizing the Non-visual: Spatial Analysis and Interaction with Information from Text Documents, *in* N. Gershon & S. Eick (eds.), *Proceedings of Information Visualization'95*, IEEE Publications, pp.51–8.

Paper 2—Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?

E. Frøkjær, M. Hertzum & K. Hornbæk, (2000), 'Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?', *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2000)*, the Haag, the Netherlands, 1st-6th April 2000, p. 345-352

Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?

Erik Frøkjær	Morten Hertzum	Kasper Hornbæk
Dept. of Computing	Centre for Human-Machine Interaction	Dept. of Computing
University of Copenhagen	Risø National Laboratory	University of Copenhagen
Copenhagen Ø, Denmark	Roskilde, Denmark	Copenhagen Ø, Denmark
+45 3532 1456	+45 4677 5145	+45 3532 1452
erikf@diku.dk	morten.hertzum@risoe.dk	kash@diku.dk

ABSTRACT
Usability comprises the aspects effectiveness, efficiency, and satisfaction. The correlations between these aspects are not well understood for complex tasks. We present data from an experiment where 87 subjects solved 20 information retrieval tasks concerning programming problems. The correlation between efficiency, as indicated by task completion time, and effectiveness, as indicated by quality of solution, was negligible. Generally, the correlations among the usability aspects depend in a complex way on the application domain, the user's experience, and the use context. Going through three years of CHI Proceedings, we find that 11 out of 19 experimental studies involving complex tasks account for only one or two aspects of usability. When these studies make claims concerning overall usability, they rely on risky assumptions about correlations between usability aspects. Unless domain specific studies suggest otherwise, effectiveness, efficiency, and satisfaction should be considered independent aspect of usability and all be included in usability testing.

Keywords
Usability measures, effectiveness, efficiency, satisfaction, information retrieval, usability testing, user studies

INTRODUCTION
Although the importance of usability is gaining widespread recognition, considerable confusion exists over the actual meaning of the term. Sometimes usability is defined quite narrowly and distinguished from, for example, utility [11], on other occasions usability is defined as a broad concept synonymous to quality in use [2]. We adopt ISO's broad definition of usability [7] as consisting of three distinct aspects:

- *Effectiveness*, which is the accuracy and completeness with which users achieve certain goals. Indicators of effectiveness include quality of solution and error rates. In this study, we use quality of solution as the primary indicator of effectiveness, i.e. a measure of the outcome of the user's interaction with the system.
- *Efficiency*, which is the relation between (1) the accuracy and completeness with which users achieve certain goals and (2) the resources expended in achieving them. Indicators of efficiency include task completion time and learning time. In this study, we use task completion time as the primary indicator of efficiency.
- *Satisfaction*, which is the users' comfort with and positive attitudes towards the use of the system. Users' satisfaction can be measured by attitude rating scales such as SUMI [8]. In this study, we use preference as the primary indicator of satisfaction.

While it is tempting to assume simple, general relations between effectiveness, efficiency, and satisfaction, any relations between them seem to depend on a range of issues such as application domain, use context, user experience, and task complexity. For routine tasks good performance depends on the efficient, well-trained execution of a sequence of actions which is known to yield stable, high-quality results [3]. For such tasks high-quality results are routinely achieved, and task completion time may therefore be used as an indicator of overall usability. For non-routine, i.e. complex tasks, there is no preconceived route to high-quality results, and good performance is primarily dependent on conceiving a viable way of solving the task [9, 14]. The efficient execution of the sequence of actions is of secondary importance. Consequently, efficient execution of the actions may or may not lead to high-quality results, and diligence is not even guaranteed to lead to task completion. This suggests that, at least for complex tasks, efficiency measures are useless as indicators of usability unless effectiveness is controlled.

Nielsen & Levy [12] analyzed the relation between efficiency and user preference in 113 cases extracted from 57 HCI studies. Their general finding was that preference predicts efficiency quite well. However, in 25% of the cases the users did not prefer the system they were more efficient in using. The ambition of finding a simple, general relationship between efficiency and satisfaction is therefore questionable [see also 1]. Studies of, for example, specific application domains may yield more precise and informative models. With respect to the relationship between satisfaction and effectiveness, Nielsen & Levy [12] note that their very comprehensive literature survey did not encounter a single study that compared indicators of these two aspects of usability.

In this paper we investigate the connection between efficiency, indicated by task completion time, and effectiveness, indicated by quality of solution. This is done by reanalyzing data from the TeSS-experiment [6] where 87 subjects solved a number of information retrieval tasks, using four different modes of the TeSS system and programming manuals in hard copy. In analyzing the data we look for correlations between efficiency and effectiveness across retrieval modes, tasks, and individual subjects.

The purpose of this paper is to emphasize the importance of accounting for all three aspects of usability in studies that assess system usability, for example to compare the usability of different designs. Effectiveness is often difficult to measure in a robust way. This may be the reason why several studies involving complex tasks refrain from accounting for effectiveness and settle for measures of the efficiency of the interaction process [for example, 5, 13]. These studies rest on the assumption that an efficient interaction process indicates that the user also performed well in terms of crucial effectiveness indicators such as solution quality. The TeSS-experiment illustrates that this assumption is not warranted—unless it can be supported by an argument that effectiveness is controlled.

The first two sections present the method and results from the TeSS-experiment, establishing the argument that efficiency and effectiveness are weakly—if at all—correlated. Next, we discuss the general relationship between the three aspects of usability, exemplifying the impact of our findings by studies from the CHI Proceedings of the years 1997-99. We then discuss the implications of our findings with regard to the selection of usability measures. In the final section, we outline our main conclusions concerning the weak and context-dependent relation between the usability aspects.

THE TESS-EXPERIMENT

The purpose of the TeSS-experiment was to compare the usage effectiveness of browsing and different forms of querying in information retrieval tasks concerning programming problems. Further, the experiment aimed at establishing a detailed description of the subjects' interaction with the TeSS system.

Experimental Conditions

To solve the tasks the subjects needed information concerning the development of graphical user interfaces in the X Window System. Access to the necessary documentation (approximately 3 Mb of text) was provided through an experimental text retrieval system called TeSS and by means of manuals in hard copy. TeSS can be operated in four different modes, each providing the user with a different set of retrieval facilities. Thus, the experiment involves five retrieval modes:

- BROWSE. In TeSS, browsing can be done by expanding and collapsing entries in the table of contents and by searching the table of contents for specific strings. The text itself is presented in separate windows.
- LOGICAL. A mode of TeSS offering conventional Boolean retrieval where queries are logical expressions built of query terms, ANDs, ORs, NOTs, parentheses, and wildcards.
- VENN. In this mode of TeSS queries are expressed by means of a Venn diagram which replaces Boolean operators with a, supposedly, more immediately understandable graphical image of intersecting sets.
- ALL. The whole of TeSS offering the combination of BROWSE, LOGICAL, and VENN.
- PAPER. In this mode searching is done in hard copies of the programming manuals, i.e. independently of TeSS.

Subjects

The subjects were 87 students in their third year of a bachelor degree in computer science. While the project was a mandatory part of the students' education, participation in the experiment by allowing the data collection to take place was voluntary and anonymous. The subjects were first-time users of TeSS and had no prior knowledge of the programming tools on which the tasks were based.

Tasks

In the TeSS-experiment each subject solved 20 information retrieval tasks. As preparation, the subject completed two practice tasks. The 20 tasks concerned whether and how certain interface properties could be achieved in a graphical user interface. To answer the tasks the subjects had to identify the relevant user interface objects, e.g. widgets, methods, and resources, and outline an implementation. As the subjects were unfamiliar with the X Window System, the tasks involved a substantial element of learning in addition to the need for retrieving specific pieces of information. Some tasks were formulated in the context of the X Window System in general; others took the user interface of TeSS as their point of departure. Two examples of tasks used in the TeSS-experiment are:

Task 5. Radio buttons are used in situations where exactly one option must be chosen from a group of options. Which widget class is used to implement radio buttons?

Task 11. The caption on the button "done" should be changed to "quit". How is that done?

Procedure

The experiment was explained to the subjects at a lecture, after which the subjects had ten days to complete the tasks. The subjects received a manual for TeSS and a two-page walk-up-and-use introduction. The system itself was available on terminals to which students have access 24 hours a day. The manual searching was done in the library where one of the authors was present three hours a day to hand out tasks and receive solutions. Upon entering the library, the subjects received hard copies of the three manuals, a sheet with the proper task, and a log sheet with fields for starting time, finishing time, and solution.

The experiment employed a within-groups design where all subjects solved the tasks in the same sequence and each subject was required to use all retrieval modes. To avoid order effects, the subjects were exposed to the retrieval modes in a systematically varied order. The 20 information retrieval tasks were clustered into five blocks. The first block was solved with one of the five retrieval modes, the second block with one of the remaining four retrieval modes. Thus the permutations of the modes on the two first blocks divided the subjects into 20 groups. The number of subjects did not allow all 5! sequences of the five modes to be included, and the 20 groups were not divided further. Rather, the order of the three remaining modes was kept the same within each group.

Data Collection and Analysis

The data collected in the experiment include a detailed log of the subjects' interaction with TeSS. The interaction log gives a time-stamped account of the commands executed by the subjects. It also includes task demarcation and solutions reached, both obtained from a separate module governing the subjects' access to TeSS. This Task Handling Module makes it possible to let the subjects work unsupervised while at the same time enforcing a strict experimental procedure. The Task Handling Module presents the tasks to the subject one at a time, gives access to the retrieval mode to be used by that subject when solving that particular task, and records his or her solution. For the PAPER retrieval mode, the subjects recorded their starting time, finishing time, and task solution on the log sheets.

The 87 subjects received 20 information retrieval tasks each, giving a potential total of 1740 answers. However, 113 answers were not submitted; 19 were excluded because they included a more than one hour long period with no logged user activity; 17 were excluded due to technical problems with TeSS; 14 were excluded because it was impossible to judge the quality of the answer; and 2 were excluded because they were solved poorly in less than two minutes, i.e., without any attempt to reach a solution. Finally, 4 subjects were excluded because they clearly did not take the experiment seriously. Thus, 11% of the answers were not submitted or excluded. The analysis is based on the remaining 1555 answers, the results of 648 hours of work performed by 83 subjects.

Grade	Mnemonic	Description
1	Very low	Failure, a completely wrong answer
2	Low	Inadequate or partially wrong answer
3	Medium	Reasonable but incomplete answer
4	High	Good and adequate answer
5	Very high	Brilliant answer

Table 1—The five-point scale used to grade the tasks

In this paper we focus on two aspects of the usability of TeSS:

- Efficiency measured as task completion time, which is extracted from the interaction log or the log sheets.
- Effectiveness measured as the quality of the solution, which was assessed by one of the authors and expressed by a grade on a five-point scale, see Table 1. As an example, a medium and a high quality solution to task 5 (see above) must identify toggle widgets as the relevant widget class. A brilliant answer also explains the use of radio groups to cluster the toggle widgets.

The following analysis is restricted to the 20 information retrieval tasks—the bulk of our data. Data concerning user satisfaction, measured as subjects' preference for one or the other retrieval mode, were collected for three implementation tasks, which followed the information retrieval tasks. The preference data show that the subjects did not prefer the retrieval mode with which they performed best. Rather, they overwhelmingly preferred ALL, the retrieval mode where they did not exclude themselves from any of the search facilities available in BROWSE, BOOLEAN, or VENN [6]. This suggests that user satisfaction is not simply correlated with performance measures such as task completion time and grade. Thus, the TeSS-experiment was another exception to the general finding of Nielsen & Levy [12] that users prefer the objectively best system.

RESULTS OF THE TESS-EXPERIMENT

Table 2 shows the relation between task completion time and grade for the 1555 tasks solved in the TeSS-experiment. A contingency analysis of this table suggests that task completion time and grade are not independent ($\chi^2[16, N=1555]=47.81, p<0.001$).

Task completion time for subjects receiving a certain grade varies much, as can be seen from the large standard deviations in Table 2. An analysis of variance shows

Task completion time / Grade (no. of observations)	$<P_{20}$	P_{20}-P_{40}	P_{40}-P_{60}	P_{60}-P_{80}	$>P_{80}$	Mean time for grade (SD)
5 (N=147)	17	35	33	31	31	24.27 (20.62)
4 (N=566)	170	121	92	96	87	21.71 (38.80)
3 (N=216)	37	48	55	38	38	24.70 (26.18)
2 (N=192)	29	35	46	36	46	26.72 (32.60)
1 (N=434)	58	72	85	110	109	28.94 (27.35)
Median grade (P_{25}-P_{75})	4 (2-4)	4 (2-4)	3 (1-4)	3 (1-4)	3 (1-4)	

Table 2—Distribution of task completion time and grade for all tasks in the TeSS-experiment (N=1555). The column to the left shows the five grades given to the tasks, cf. Table 1. The next columns show the number of tasks in each of five intervals based on the 20, 40, 60, and 80 percentiles of task completion time. The rightmost column shows the mean time in minutes for a certain grade and, in parentheses, the standard deviation. The bottom row shows the median grade for each time interval, indicating the variation in grades by the 25- and 75-percentile.

significant variation in task completion times between different grades ($F[4,1550]=3.31$, $p<0.01$). However, we did not find any pairwise differences between grades using Tukey's post hoc test at a five-percent significance level.

The tasks in any of the five intervals of task completion times shown in Table 2 received markedly different grades. Between time intervals there is significant variation in grades (analysis of variance with time interval as the independent and grade as the dependent variables, $F[4,1550]=9.10$, $p<0.001$). Pairwise comparisons of the five time intervals using Tukey's post hoc test show that the 20% fastest solved task receive significantly higher grades than the 60% slowest solved tasks. Similarly, solutions to tasks in the P_{20}-P_{40} time interval receive significantly higher grades than solutions in the time intervals P_{60}-P_{80} and $>P_{80}$.

Spearman's rank order correlation analysis shows that task completion time and grade are significantly correlated in tasks solved in the TeSS-experiment ($r_s=-0.156$, two-tailed p-level <0.001). Using more time for completing a task is thus correlated with receiving a lower grade. However, the correlation between time and grade is weak; only two percent of the variation in grade can be predicted from task completion time ($r_s^2=0.024$). According to [4] a correlation of this magnitude is negligible.

Retrieval mode (no. of observations)	Mean time (SD)	Median grade (P_{25}-P_{75})	r_s	p	r_s^2%
Browse (N=310)	22.88 (20.89)	3 (1-4)	-0.150	0.008	2.2
Logical (N=307)	30.15 (34.70)	3 (1-4)	-0.089	0.119	-
Venn (N=305)	25.79 (25.45)	3 (1-4)	-0.107	0.062	-
All (N=314)	30.80 (51.84)	3 (1-4)	-0.128	0.030	1.6
Paper (N=319)	15.66 (11.27)	4 (2-4)	-0.265	0.001	7.0

Table 3—Correlation between time and grade in different retrieval modes. The first column shows the retrieval modes, and the second and third columns the mean time in minutes and the median grade for each mode. Columns four to six show the Spearman correlation coefficient between time and grade r_s, the significance level for the correlation p, and the strength of the correlations at a five-percent significance level r_s^2%.

To control for interplay between the design of the experiment and the weak correlation found, we performed a partial correlation analysis of the TeSS data. In the partial correlation analysis, the influence from different tasks and retrieval modes is removed from the correlation coefficient between time and grade [4]. This analysis also reveals a weak but statistically significant correlation between task completion time and grade (Spearman's partial correlation coefficient r_s[time,grade| configuration,task]=-0.170, $p<0.001$).

These analyses show that at the general level efficiency and effectiveness are only weakly correlated. In spite of this, time and grade could be correlated at a more detailed level of analysis, hereby undermining the conclusion at the general level. In the following sections we therefore analyze whether time and grade are correlated for specific retrieval modes, tasks, or subjects.

Correlation between Time and Grade for Different Retrieval Modes

The retrieval modes LOGICAL and VENN—the only retrieval modes requiring the subjects to formulate queries—do not show a significant correlation between time and grade (see Table 3). The retrieval modes BROWSE, ALL, and PAPER all show a statistically significant but weak correlation between task completion time and grade (r_s^2% between 1.6 and 7.0). The tasks solved in the retrieval mode PAPER have a numerically larger correlation between time and grade than the other retrieval modes. However, the correlation for PAPER is still weak and not significantly different from the correlations for BROWSE and ALL (Fisher's r-to-z

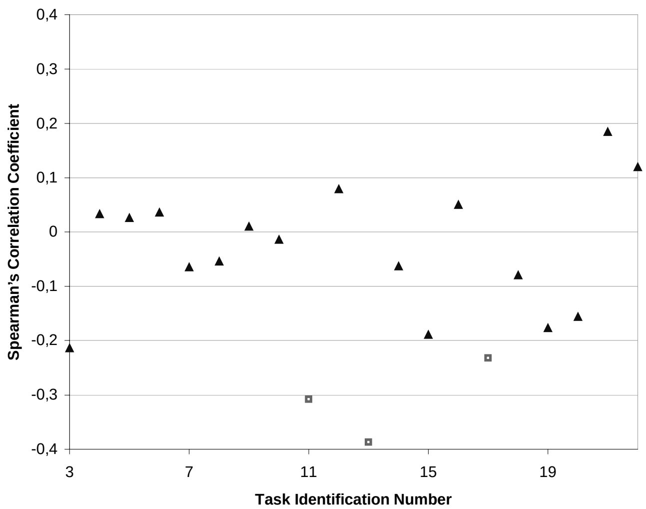


Figure 1—Correlation between time and grade for different tasks. The figure shows Spearman's correlation coefficient (r_s) for each of the 20 information retrieval tasks. Each task has been solved by between 69 and 81 subjects. Time and grade are significantly correlated for tasks 11, 13, and 17. These tasks appear as squares in the figure. The task identification numbers begin at 3, because tasks 1 and 2 are tasks used for training [6].

transformation, ALL vs. PAPER: z=-1.783, p>0.075, BROWSE vs. PAPER, z=-1.504, p>0.133).

Correlation between Time and Grade for Different Tasks

The correlation between task completion time and grade varies somewhat across the tasks (see Figure 1). For 85% of the tasks there is no correlation between time and grade. However, three tasks show a significant correlation between time and grade: task 11 (r_s=-0.308, p<0.007), task 13 (r_s=-0.387, p<0.001), and task 17 (r_s=-0.232, p<0.040). For these tasks between 5% and 15% of the variation in grade can be predicted from time, where more time spent is correlated with lower grade.

Task 11 and task 13 have a higher average grade than the other tasks (task 11: mean grade 3.42, t[1393]=-3.734, p<0.001; task 13: mean grade 3.72, t[1398]=-5.739, p<0.001). Task 13 is also solved faster than the other tasks (mean completion time 13.43 minutes, t[1398]=3.316, p<0.001). The description of these tasks given to the subjects specifies in detail some of the central interface objects of the tasks (see for example the wording of task 11 showed earlier). For task 17 it is only the relation between time and grade that is significant, individually neither time nor grade differs significantly from the other tasks.

Correlation between Time and Grade for Different Subjects

Looking at the average performance of subjects, the tasks solved by 12 of the subjects show a significant correlation between time and grade (see Figure 2). These correlation

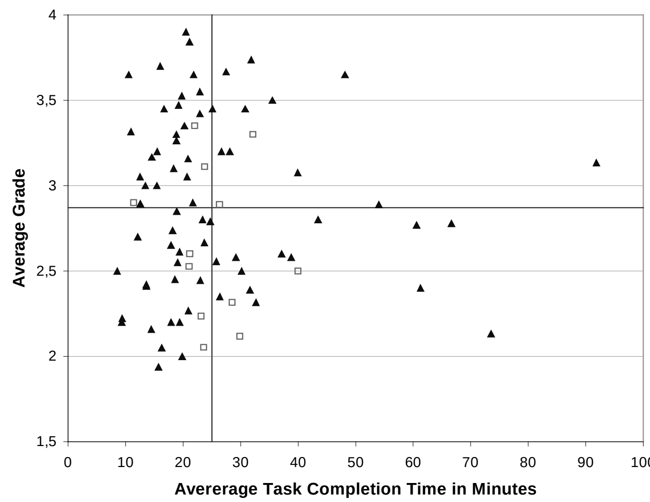


Figure 2—Average time and grade for each of the 83 subjects included in the data analysis. The horizontal line indicates the overall mean grade (2.87), the vertical line the overall mean time (25 min.). Subjects with a significant correlation between time and grade appear as squares, other subjects appear as triangles.

coefficients are all negative, suggesting that more time spent is correlated with lower grade (r_s between –0.758 and –0.453). For 86% of the subjects, time does not predict grade at all.

It is difficult to find a common denominator for the subjects where time and grade are correlated. The average time and grade of those subjects vary above and below the mean time and grade for subjects (see Figure 2). However, there is a significant difference between the grade for subjects with a significant correlation between time and grade and those without (Wilcoxon test, z=2.393, p<0.017). Subjects who obtain a correlation between time and grade did not use a specific retrieval mode for certain tasks (Chi-square test of which retrieval mode was first used, χ^2[4, N=12]=3.833, p>0.05).

Summary of Correlations between Usability Measures

Our analysis of the TeSS-experiment shows that efficiency (measured as task completion time) and effectiveness (measured as grade) are either not correlated or correlated so weakly that the correlation is negligible for all practical purposes. For the individual retrieval modes, a weak correlation is found for three of the modes, while two of the modes do not show any significant correlation between task completion time and grade. Task completion time and grade are not correlated for 85% of the tasks. Finally, only 14% of the subjects display a significant correlation between time and grade—for the large majority no correlation is found. These results and the previous results [6] concerning satisfaction and effectiveness (cf. the section Data Collection and Analysis, last paragraph) show that assumptions about correlations between effectiveness,

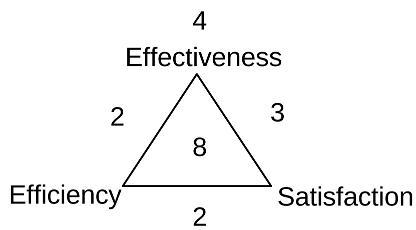


Figure 3—The usability aspects measured in the 19 studies of complex tasks from CHI '97 to CHI '99. Eight of these CHI-studies include measures of all three usability aspects, seven CHI-studies measure two aspects, and four CHI-studies only one aspect.

efficiency, and satisfaction do not seem to hold in the context of the TeSS-experiment.

CORRELATIONS BETWEEN ASPECTS OF USABILITY

We now extend the discussion of correlations between aspects of usability by including studies of computer support for complex tasks published in the CHI Proceedings for the years 1997-99. A total of 19 studies investigate aspects of usability in sufficient detail to enable an analysis of their choice of usability measures, see Figure 3. Eight (42%) of the 19 studies cover all three usability aspects. The other 11 studies, implicitly or explicitly, rely on assumptions of correlations between the different usability aspects, or seem confident that their choice of only one or two aspects of usability is sufficient to capture overall usability.

The only CHI-study with an analysis of correlations between the three aspects of usability

Of the eight studies including measures of all three usability aspects, only the study by Walker et al. [17] has analyzed the correlations between the aspects. Let us summarize their study, so the reader can see that the correlation analysis pays off.

Walker et al. compare two different designs of a spoken language interface to email: (a) a mixed-initiative dialogue, where the users can flexibly control the dialogue, and (b) a system-initiative dialogue, where the system controls the dialogue. The study measures effectiveness by qualitative measures such as automatic speech recognition rejects, efficiency by number of dialogue turns and task completion time, and user satisfaction by a multiple-choice survey. The results show that even though the mixed-initiative dialogue is more efficient, as measured by task completion time and number of turns, users prefer the system-initiative dialogue.

A correlation analysis with user satisfaction as the dependent variable uncovers how "…users' preferences are not determined by efficiency per se, as has been commonly assumed. One interpretation of our results is that users are more attuned to qualitative aspects of the interaction." [17, p. 587]. The number of automatic speech recognition rejects contributes the most to user satisfaction. Walker et al. suggest that the users' preference for the system-initiative dialogue arises from it being easier to learn and more predictable. This result was contrary to the authors' initial hypothesis and illustrates the importance of measuring efficiency, effectiveness, and satisfaction independently, as opposed to basing conclusions about one of them on measures of the others.

Two CHI-studies without any measure of effectiveness

Two CHI-studies concerning computer support for complex tasks, entitled "Time-compression: systems concerns, usage, and benefits" [13] and "Effects of awareness support on groupware usability" [5], do not include any measure of the quality of the outcome of the users' interaction with the system. Below we comment on these two studies, and show how their conclusions about overall usability are jeopardized by their incomplete choice of usability measures.

In the first study, Omoigui et al. [13] analyze how time-compression can be used to enable quick video browsing. An experimental time-compression system was used for comparing different granularities of the time-compression (discrete vs. continuous) and differences in the latency (long wait-time vs. no wait-time) experienced by users after adjusting the degree of time-compression. Omoigui et al. measure efficiency by savings in task time and the use of time-compression, and they measure satisfaction by, e.g., user feedback and preference indicated by usage of time-compression during video browse sessions. As already mentioned, no effectiveness measures were employed, although effectiveness could have been measured as the accuracy and completeness of the subjects' verbal summary of each video. In the concluding remarks, Omoigui et al. emphasize efficiency as the important aspect of time-compression systems: "Quite surprisingly though, there are no significant differences in the time-savings under the three conditions. Thus the implementers are free to choose the simplest solution..." [13, p. 142]. This conclusion neglects the satisfaction measures, which indicate that real differences might exist between the experimental conditions: "… several subjects commented in post-study debriefing that the long latency and discrete granularity conditions had affected their use of the time compression feature. The subjects felt that they made fewer adjustments and watched at a lower compression rate when long latency and discrete granularity were used." [13, p. 141]. An analysis of the correlations between the efficiency and satisfaction measures might have shed further light on the differences between conditions, as might solid measures of effectiveness.

In the second study, Gutwin and Greenberg [5] analyze whether enhanced support for workspace awareness improves collaboration. In an experiment, they compare users' performance on two real-time groupware systems where workspace miniatures were used to support workspace awareness. The basic miniature shows information only about the local user, the enhanced miniature about others in the workspace as well. Efficiency is measured by task completion time and communication efficiency; satisfaction is measured as preference for one or the other system. The correlations between the measures are not analyzed, and no measure of effectiveness is employed. The overall conclusion of the study is that workspace-awareness information reduces task completion time, and increases communicative efficiency and user satisfaction. The support for this conclusion is weak. For one out of the three task types, task completion time was not reduced. For two task types out of the three, the communicative efficiency was not increased. All 38 participants preferred the awareness-enhanced system, suggesting that the employed measures of usability are incomplete: "The overwhelming preference for the interface with the added awareness information also suggests that there were real differences in the experience of using the system, but that our measures were insensitive to these differences." [5, p. 517]. These differences might have been more explainable if the study had included measures of effectiveness, making possible an analysis of how users' preferences were affected by the quality of the outcome of their activities.

SELECTION OF USABILITY MEASURES

We believe that the weak correlation between effectiveness, efficiency, and satisfaction has three implications regarding the choice of measures in evaluations of system usability.

First, it is in general recommendable to measure efficiency, effectiveness as well as satisfaction. When researchers or developers use a narrower selection of usability measures for evaluating a system they either (a) make some implicit or explicit assumptions about relations between usability measures in the specific context, or (b) run the risk of ignoring important aspects of usability. In our analysis of the CHI-studies we have shown how interpretation of experimental data based on only one or two usability aspects leads to unreliable conclusions about overall usability. Given that the three usability aspects capture different constituents of usability—we have not seen arguments to the contrary for complex tasks—there is no substitute for including all three aspects in usability evaluations.

Second, at the moment no clear-cut advice can be given about which usability measures to use in a particular situation. On the contrary, identifying the usability measures that are critical in the particular situation should be recognized as a central part of any evaluation of system usability. This requires a firm understanding of how tasks, users, and technology interact in constituting the use situations within the particular application domain [10, 16]. The study by Su [15] is an illustrative example of the kind of work needed to distinguish and refine performance measures. Su investigated the correlation between 20 measures of information retrieval performance in an academic setting, and suggests a best single measure (the user's perception of the value of the search result as a whole) and best pairs of measures of information retrieval performance. Such work may lead to the development of reliable, domain-specific collections of critical performance measures. General descriptions of the relation between usability aspects [e.g. 12] will not aid the selection of usability measures, since there is no way of knowing in advance whether efficiency, effectiveness, and satisfaction are actually correlated in a particular situation.

Third, effectiveness measures oriented toward the outcome of the user's interaction with the system are gaining attention in usability evaluation [2], although two of the CHI-studies discussed earlier did not include such measures. The development of valid and reliable outcome measures is a prerequisite for assessing overall system usability and is necessary for working systematically with improving the usability of systems supporting users in solving complex tasks.

CONCLUSION

The relations between efficiency, effectiveness, and satisfaction—the three aspects of usability—are not well understood. We have analyzed data from a study of information retrieval and found only a weak correlation between measures of the three usability aspects. Other studies imply that for complex tasks in other domains, a similarly weak correlation between usability measures is to be expected. In general, we suggest that efficiency, effectiveness, and satisfaction should be considered independent aspects of usability, unless domain specific studies suggest otherwise.

Studies that employ measures of only a subset of the three usability aspects assume either that this subset is sufficient as an indicator of overall usability or that the selected measures are correlated with measures covering the other aspects of usability. As we have exemplified with an analysis of studies from previous CHI Proceedings, such assumptions are often unsupported. Hence, these studies jump to conclusions regarding overall usability while measuring, say, efficiency only. This is a problem for the HCI community, since more than half of the last three years of CHI-studies concerning complex tasks do not measure all aspects of usability.

Usability testing of computer systems for complex tasks should include measures of efficiency, effectiveness, and user satisfaction. In selecting these measures, the application domain and context of use have to be taken into account so as to uncover the measures that are critical in the particular situation. Discovering solid measures of effectiveness seems especially critical.

ACKNOWLEDGEMENTS

Morten Hertzum was supported by a grant from the Danish National Research Foundation. The design and implementation of TeSS as well as the design and execution of the experiment were done in collaboration with Jette Broløs, Marta Lárusdóttir, Kristian Pilgaard, and Flemming Sørensen. We wish to thank the students who participated in the experiment as subjects, and the CHI-reviewers. We are indebted to Per Settergren Sørensen for his support on statistical issues and to Peter Naur for many judicious proposals for clarification.

REFERENCES

1. Bailey, R.W. Performance vs. preference, in *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting* (Seattle WA, October 1993), HFES, 282-285.
2. Bevan, N. Measuring usability as quality of use. *Software Quality Journal* 4 (1995), 115-150.
3. Card, S.K., Moran, T.P. and Newell, A. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM* 23, 7 (1980), 396-410.
4. Cohen, J. and Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates, Hillsdale NJ, 1975.
5. Gutwin, C. and Greenberg, S. Effects of awareness support on groupware usability, in *Proceedings of CHI '98* (Los Angeles CA, May 1998), ACM Press, 511-518.
6. Hertzum, M. and Frøkjær, E. Browsing and querying in online documentation: A study of user interfaces and the interaction process. *ACM Transactions on Computer-Human Interaction* 3, 2 (1996), 136-161.
7. ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability (1998).
8. Kirakowski, J. and Corbett, M. SUMI: The software usability measurement inventory. *British Journal of Educational Technology* 24, 3 (1993), 210-212.
9. Naur, P. Intuition in software development, in H. Ehring et al. *Formal Methods and Software Development*, Vol. 2, Lecture Notes in Computer Science 186, Springer, Berlin, 1985. Also in Naur, P. *Computing: A Human Activity*. ACM Press/Addison-Wesley, New York, 1992, 449-466.
10. Newman, W. and Taylor, A. Toward a methodology employing critical parameters to deliver performance improvements in interactive systems, in *Proceedings of INTERACT '99* (Edinburgh, August 1999), IOS Press, 605-612.
11. Nielsen, J. *Usability Engineering*, Academic Press, Boston, 1993.
12. Nielsen, J. and Levy, J. Measuring usability: Preference vs. performance, *Communications of the ACM* 37, 4 (1994), 66-75.
13. Omoigui, N., He, L., Gupta, A., Grudin, J. and Sanocki, E. Time-compression: Systems concerns, usage, and benefits, in *Proceedings of CHI '99* (Pittsburg PA, May 1999), ACM Press, 136-143.
14. Rasmussen, J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13, 3 (1983), 257-266.
15. Su, L.T. Evaluation measures for interactive information retrieval. *Information Processing & Management* 28, 4 (1992), 503-516.
16. Van Welie, M., van der Veer, G.C. and Eliëns, A. Breaking Down Usability, in *Proceedings of INTERACT '99* (Edinburgh, August 1999), IOS Press, 613-620.
17. Walker, M.A., Fromer, J., Di Fabbrizio, G., Mestel, C. and Hindle, D. What can I say?: Evaluating a spoken language interface to email, in *Proceedings of CHI '98* (Los Angeles CA, May 1998), ACM Press, 582-589.

Paper 3—Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces

K. Hornbæk & E. Frøkjær, (2001), 'Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces', *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2001)*, Seattle, WA, 31st March-5th April 2001, p. 293-300.

Reading of Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces

Kasper Hornbæk
Department of Computing
University of Copenhagen
Copenhagen East, Denmark
+45 35321452
kash@diku.dk

Erik Frøkjær
Datalogi
Roskilde Universitetscenter
Roskilde, Denmark
+45 46743858
erikf@ruc.dk

ABSTRACT
Reading of electronic documents is becoming increasingly important as more information is disseminated electronically. We present an experiment that compares the usability of a linear, a fisheye, and an overview+detail interface for electronic documents. Using these interfaces, 20 subjects wrote essays and answered questions about scientific documents. Essays written using the overview+detail interface received higher grades, while subjects using the fisheye interface read documents faster. However, subjects used more time to answer questions with the overview+detail interface. All but one subject preferred the overview+detail interface. The most common interface in practical use, the linear interface, is found to be inferior to the fisheye and overview+detail interfaces regarding most aspects of usability. We recommend using overview+detail interfaces for electronic documents, while fisheye interfaces mainly should be considered for time-critical tasks.

Keywords
Reading activity, electronic documents, information visualization, user study, usability, information retrieval

INTRODUCTION

We investigate if interfaces using information visualization techniques can support reading of electronic documents. Although several interfaces for electronic documents using information visualization have been proposed, little is known about the usability of such interfaces. In an experiment, we compare 20 subjects' reading activity in a linear, a fisheye, and an overview+detail interface. We describe differences in usability between the three interfaces, describe different patterns of reading between interfaces, and illuminate some individual differences in reading. Based on these differences, we offer advice to designers of electronic documents regarding the usability of linear, fisheye, and overview+detail interfaces.

Our focus on reading of electronic documents has two motivations. First, electronic documents are increasingly being used in professional activities and are widely read on the World Wide Web, in online journals, and in electronic newspapers. Sellen & Harper [27], describing the use of paper and electronic documents among analysts at the International Monetary Fund, assess that 14% of the time analysts worked with documents, they used electronic documents only. Analysts used a combination of paper and electronic documents 35% of the time. A study of World Wide Web usage [7] found that users spend at least twice as much time using the information they find, compared to searching, browsing, or any other activity. Reading is the main activity in using information. A study of the usage of electronic journals [29] reports that 28% of a sample of 75 academics used such journals—mainly because of the accessibility of the journals and because the academics could read such journals at their desktop. Hence, improved support of reading represents an important challenge to interface designers with an impact on a range of activities and a large group of users.

Our second motivation stems from the belief that reading play a central role in information access and use. When users access a collection of electronic documents, they most often face a problem that they believe can be resolved by information in the collection [1,20]. Although gaining an overview of the collection and formulating queries are important activities, the problematic situation that motivated users to access the collection is ultimately resolved through interacting with the documents [1]. Users' interaction with documents are both physical—such as navigating to certain sections—and mental—such as trying to grasp the intention of the author with a particular sentence or to integrate the information in the document with their own ideas. Interacting with and reading documents are thus necessary for successfully resolving the users' problems. Much research has tried to improve users information access and use by better search engines, support for query construction, or collection overviews [8,28]. Here we take a complementary approach, focusing on the reading of individual electronic documents.

The remainder of this paper is structured as follows. First, we sketch related research on developing more usable electronic documents, focusing on the use of information visualization techniques. Then, we present an experiment comparing the usability of a linear, a fisheye, and an overview+detail interface used for reading scientific papers. Finally, we discuss limits and benefits of the overview+detail and fisheye interfaces, and draw some implications for design of information access systems and electronic documents.

RELATED RESEARCH

The problems users face when reading electronic documents are well described, as are ways to improve the readability and navigability of such documents (see [11,21,22,26] for overviews). Here we briefly review previous attempts to use information visualization techniques for presenting electronic documents.

One group of interfaces for electronic documents shows a graphical overview of the document separated from the detailed content of the document [4,6,13,16] (See [24] for a general discussion). Seesoft [13] maps source code into an overview by letting one line of code correspond to a thin row in the overview, color-coded to display useful information about the program. In the Thumbar [16], a graphical overview of World Wide Web pages is shown next to the display of the page itself. Concepts in the user's profile are highlighted both on the overview and on the web page. Byrd [6] extends scrollbars for an interface that presents electronic documents so that the distribution of query terms in the document is shown on the scrollbar using color-coding. This extension is believed to support navigation in a document and to aid users in gaining an overview of the distribution of query terms within the document. Boguraev et al. [4] present automatically generated summaries of electronic documents together with an overview of the entire document. The user can use the summaries to access the detailed content of the document. While we know of no empirical evaluations of graphical overview+detail interfaces for electronic documents, studies of text overviews for electronic documents and graphical overviews of hypertext suggest that overviews might be effective [9,10]. Note also the important Superbook studies [12], which showed that an expandable table of contents and a word lookup function improved performance by 25% over searching in a paper manual.

Several attempts have been made at distorting parts of the document [17,18,23,25]. The aim of the distortion is to show the entire document at once or to make the salient parts of the document visible. In the Document Lens interface [25], all pages in a document are shown laid out in rows. The user can zoom in on pages to make them readable using a rectangular focus, and pan making other pages come into focus. The pages not in focus are distorted to fit the area outside of the rectangular focus. Flip zooming [17] uses a similar layout of pages, but can show pages out of focus as a heading at readable size, rather than distorting them. The fisheye view [15] shows only those parts of a document that has a degree of interest above a certain threshold. The degree of interest for a part of the document is calculated from an a priori measure of importance, e.g. the part being a headline, as well as distance between the part and the current point of view. Kaugars [18] describe a system that presents electronic documents in four increasingly informative ways, one of which focus on the first couple of paragraphs that contain query terms. The rest of the document is distorted to fit the remaining part of the window. Páez et al. [23] present an interface for electronic documents, where the font size is bigger for the title, headings, and key sentences than for other parts of the document. Initially, the entire document is fitted on the screen. The user can then zoom in and read the interesting sections. Páez et al. did not find the zoomable interface for electronic document to be more effective than hypertext. In general, little is known about the usability of distorted electronic documents.

EXPERIMENT

In the experiment, we compared how subjects' reading activity was supported by a linear, a fisheye, and an overview+detail interface. Subjects answered questions about object oriented systems development and wrote essays that summarized and commented journal papers. We analyzed usability differences between the interfaces by grades given for the answers to the questions and the essays, by satisfaction and preference data, and by a log of the subjects' interactions with the interfaces.

Interfaces

Figure 1 shows screenshots of the three interfaces used in the experiment. In the linear interface, the document is shown as a linear sequence of text and pictures, similar to how documents are presented on paper and in most interfaces for electronic documents in practical use.

In the fisheye interface, certain parts of the document are considered more important than other parts; these parts are always readable. The remaining parts of the document are initially distorted below readable size, but can be expanded and made readable if the user clicks on them with the mouse. The aim of the fisheye interface is to reduce the time taken to navigate through a document and to support readers in employing an overview oriented reading style—first focusing on the important sections of the document, then expanding sections and reading the details. All sections can be expanded simultaneously, or returned to their initial state, by selecting a menu item in a pop-up menu.

Two measures are used to determine which sections to consider important. First, research in automatic summarization of documents suggests that sentences selected from the beginning and end of a document unit are among the best indicators of the content of that unit [5,19]. Hence, the first and last paragraph of a section is considered important and is initially readable; the other parts of the section are considered to be less important and are initially distorted. This scheme is recursively applied to subsections, so that when a section is expanded, only the first and last parts of subsections are readable. Second, empirical research has found that readers often attend to and find certain components of a document especially useful [3,11]. Therefore abstracts and section headings are always visible, and graphics and tables are diminished less than text. In the fisheye interface, the initial size of the documents used in the experiment was 25% of their size in the linear interface.

In the overview+detail interface, the document is shown as a linear sequence of text and pictures (the detail pane) together with a tightly coupled overview of the document (the overview pane). The position of the view of the document shown in the detail pane is indicated in the overview pane with a rectangular field-of-view. The field-of-view can be dragged to change which part of the document is shown in the detail pane. The user can also click on the overview, which changes which part of the document that is shown in the detail pane, effectively functioning as a scrollbar. The overview pane is a semantic zoom of the document, where section and subsection headings are shown at a fixed size. The remaining text and pictures in a section are zoomed to fit the space allocated to show that section, determined by the ratio between the length of that section in the detail pane, and the total length of the document. For the six documents used in the experiment, this ratio was on average 1:17. We believe that the semantic zoom and the stability of the overview pane is the main improvement over previous overview+detail interfaces for electronic documents.

For all three interfaces, the documents can be navigated using the mouse or the keyboard and have immediate feedback when scrolling. It is also possible to highlight words, which makes words in the document containing one or more of the words entered by the user appear red. Highlighted words are also shown in the overview pane and in sections in the fisheye interface that are diminished.

Design

The experiment employed a 2×3 within-subjects factorial design, with task and interface type as independent variables. The experiment consisted of three sessions, in each of which 20 subjects used one interface to solve a task of each type. Each session lasted approximately one hour and 45 minutes, for a total of 106 hours of experimental data. Tasks and interfaces were systematically varied and counterbalanced. We formed six groups based on all sequences of interfaces. The tasks for these six groups were found by randomly choosing latin squares such that the three interfaces and the three sessions have an approximately equal number of different tasks.

Subjects

The subjects in the experiment were students at the Department of Computing, University of Copenhagen (DIKU), who chose to participate in a course involving the experiment. The subjects had studied computer science for a mean time of 6.5 years. Of the 20 subjects, 15 were males and five females, with a mean age of 27. Sixteen subjects reported to use computers every day, four subjects several times a week. Fourteen subjects had self-reported

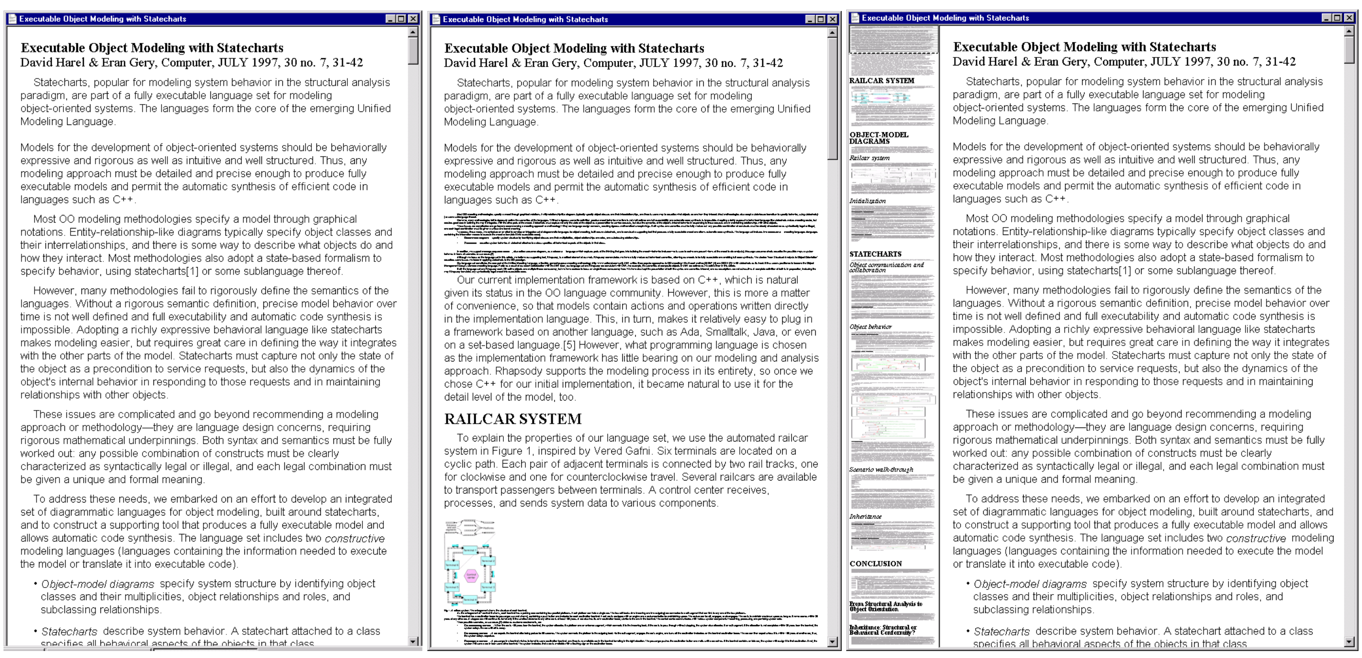


Figure 1—The linear (left), fisheye interface (middle), and overview+detail interface (right). The fisheye interface has certain parts of the document distorted below readable size. The distorted sections can be made readable by clicking on them with the mouse. The right part of the overview+detail interface is the detail pane, which is similar to the linear interface. The left part of the window is the overview pane, which shows the entire document zoomed to fit the window height. At the top of the overview pane is shown the field-of-view (dark gray area), which can be moved and dragged to change the content of the detail pane.

familiarity with object oriented systems development from courses, 11 subjects had such familiarity from systems development projects.

Tasks and Documents

The subjects were given two types of tasks: essay tasks and question-answering tasks. The essay tasks and the question-answering tasks correspond to two of what has been suggested as four typical reading tasks: reading-to-learn-to-do and reading-to-do [26]. In essay tasks, subjects read a document to learn the main content of that document. Afterwards and without access to the document, they wrote a one-page essay, stating the main theses and ideas of the document. Subjects were also requested to give approximately one page of comments about the document, which could serve as starting points for a classroom discussion. The subjects received the description of the tasks before beginning to read the document. After writing the essays, subjects were asked to answer six questions about the document just read. The subjects did not know these questions while reading the document; we therefore call these questions incidental-learning questions. Examples of incidental-learning questions include: "Which integrity problems can occur in what the author calls the simple business application architecture?" and "Which problems did the authors experience with respect to using object oriented databases?"

The second task type was question-answering tasks, where subjects answered six questions about a document, one question at a time. The six questions were varied as to 1) position in the document where the answer can be found (in the first or last part of the document); 2) how easily accessible the sentences or sections containing the answer are (whether they are near section beginnings, tables or figures); and 3) the usefulness of the words describing the question as terms for highlighting (whether or not the question contained terms that were located near the answer). Three examples of questions are: "What is, according to the paper, the biggest problem in relation to automatically transforming procedural code to object oriented code?", "What is the difference between structural and behavioral inheritance?", and "What is according to the author the difference between analysis and design?".

The documents used in the experiment were six IEEE journal papers, chosen from the top documents retrieved in response to a query on "user oriented systems development object oriented uml" in the Digital Library Initiative test bed at University of Illinois at Urbana-Champaign [2]. The paper versions of the documents were between 8 and 14 pages, contained on average four figures, and included one document with tables and one document with formulae. No subjects indicated that they previously had read any of the papers.

The descriptions of the tasks, the answers to tasks, the training material, and the satisfaction questionnaires were all in the native language of the subjects, Danish.

Dependent Measures

We measure the usability of the three interfaces by including measures of effectiveness, satisfaction, and efficiency, as recommended in [14]. Effectiveness of the interaction with the three interfaces is measured as the grade received for the answers to the tasks. The answers were graded blind by the first author, i.e., without any knowledge of which subject had made the answer or with which interface the answer had been made. We used a five point grading scale, ranging from zero—a missing or completely wrong answer—to four—an outstanding and well-substantiated answer. Table 1 shows an explanation of the grades. For the question-answering tasks, grades were given according to how many aspects of the question the answered covered. A classification of main ideas in the documents and important aspects of questions were developed to assist grading. For the incidental-learning questions, we counted the number of correct answers, resulting in a score from 0 to 6. Subjects in the experiment graded three randomly chosen sets of answers to the experimental tasks, as well as their own answers. They used the same scale for grading as the author. We wanted to use their grading as a subjective perception of the quality of the answers to the tasks.

Satisfaction was measured in three ways. After using each interface, subjects answered twelve questions about the perceived usability of the interface and their experience with solving the tasks. After having used all three interfaces, subjects indicated which they preferred. Subjects also wrote comments about the interfaces after using each of them, and described why the preferred using one of the interfaces.

The subjects' interactions with the three interfaces were logged. The main efficiency measure, time usage, is derived from the data logged. No time limit was imposed on the tasks. However, subjects were made aware of how

Grade	Meaning
0	Completely wrong or missing answer.
1	Poor or imprecise answer. The answer is incomplete, describing only one aspect of the question, or is only partially correct.
2	Average answer. The answer describes relevant aspects of the questions and is in reasonable agreement with the document. For essays tasks, the comments raise some relevant problems in the paper and are substantiated.
3	Good answer. The answer describes many relevant aspects of the document and is in complete agreement with the document. For essay tasks, the comments raise relevant questions and are well substantiated.
4	Outstanding and completely adequate answer. The answer describes all relevant aspects of the question, includes additional relevant information, and is clearly written. For the essay tasks, the comments raise important questions in a thorough and substantiated way.

Table 1—The grading scale used for grading the experimental tasks.

much time they had used when reading one paper for more than one hour, or when they took more than 30 minutes to answer one of the six questions about a document.

Procedure

The experiment took place in a room without external disturbances, where two subjects participated at a time. Upon arriving, the subjects were told about the purpose of the experiment. Next, subjects filled out a questionnaire about age, sex, their use of computers, the use of computers to read scientific documents, and their familiarity with the object oriented systems development. Then, subjects were trained in using the three interfaces until they felt confident in operating these. Training was supported by a two-page description of the specifics of operating the interfaces. The subjects also completed three training tasks, which introduced the subjects to the interfaces, and the question-answering and essay tasks. The mean time used to complete the training tasks was 35 minutes. After training, the subjects completed the first session of the experiment. Subjects returned the next day to the lab and completed the remaining two sessions.

The subjects received the tasks on sheets of paper, on which they also wrote the answers for the question-answering tasks. When subjects finished reading documents they were writing essays about, they received paper and pencil for writing the essay. The subjects were not allowed to write notes while reading the documents they wrote essays about.

Approximately four days after participating in the experiment, subjects received the documents used in the experiment, four sets of answers to the experimental tasks, including their own, and instructions on how to grade the answers. Subjects did not receive information on who had made the answers or the interface used for making the answer.

Data Analyses

Of the 20*3 possible solutions to the essay tasks, one subject did not complete a task, and one task was dropped because of a time usage three interquartile ranges above the 75-quartile, leaving 58 observations. For the question-answering tasks, out of 360 (20*3*6) possible answers, one subject failed to complete the task, leaving 354 answers. One subject's grading of one answer in a question-answering task was not done. We analyzed the data by ANOVAs with interface type, task, session, and subject as independent variables. Essay tasks and question-answering tasks were analyzed separately. All post-hoc tests were done using a Bonferroni test at a 5% significance level.

RESULTS

The results are divided into questions of how effectively subjects read documents, the subjects' satisfaction, and the subjects' efficiency. We also describe some differences in how documents are read in the three interfaces.

Effectiveness—Grades and Incidental Learning

The effectiveness measures are summarized in Table 2. Using the author's grading of the 58 essay tasks, we find a significant influence of interface on the grade obtained, $F[2,32]=4.16$, $p<.05$. A Bonferroni post-hoc test shows a significant difference at the 5% level between the overview+detail and the two other interfaces, suggesting that essays written after reading documents with the overview+detail interface receive higher grades. We find no significant difference between interfaces using the subjects' own grading of the essay tasks, $F[2,33]=.473$, $p>.6$.

The number of correctly answered incidental-learning questions is significantly different between the three interfaces, $F[2,32]=6.804$, $p<.005$. A post-hoc test shows that subjects using the fisheye presentation answered significantly fewer incidental-learning questions than subjects using the linear and overview+detail interface. Subjects using the fisheye interface answered on average 0.78 and 1.16 fewer questions than subjects using the linear and overview+detail interface, respectively.

For the question-answering tasks, no influence from interface was found on subjects' grading, $F[2,312]=.121$, $p>.88$, or on the author's grading, $F[2,313]=.179$, $p>.83$.

Satisfaction

Nineteen of the subjects prefer using the overview+detail interface; one subject prefers the linear interface. In their motivation for preferring the overview+detail interface, 10 subjects mention the overview of the documents structure and titles as an important reason; six subjects mention that the overview+detail interface support easy navigation.

Table 3 shows the subjects' answers to the questionnaires filled out after using each of the interfaces. We compared the answers using paired t-tests with a Bonferroni-

Interface	Essay tasks (N=58)			Question-answering tasks (N=354)	
	Author's grading	Subjects' grading	No. correct incidental-learning questions	Author's grading	Subjects' grading
Linear	2.00 (.86) **-**	2.35 (.75)	4.20 (1.24) +	1.99 (.94)	2.63 (.93)
Fisheye	1.95 (.78) **-**	2.32 (.67)	3.42 (1.22) -	2.04(1.04)	2.68 (.91)
Overview+Detail	2.47 (.84) **+**	2.53 (.61)	4.58 (1.22) +	2.08 (1.03)	2.66 (.95)

Table 2—Effectiveness of the three interfaces. The table shows the first authors grading of the experimental tasks, the subjects own grading, and the number of correct answers to incidental learning questions. Standard deviation is given in parentheses. A plus indicate a significant difference at a 5% significance level to the interfaces marked with minus.

adjustment of 0.05/12*3≈.0013. The overview+detail interface is preferred to the two other interfaces overall, as well on the dimensions terrible-wonderful, and frustrating-pleasant. Subjects score the fisheye interface significantly lower on the dimension confusing-clear than the overview+detail interface. Subjects also score the overview+detail interface higher compared to the linear interface on the question whether the documents were easy or hard to overview. Note, that this question is not as leading in Danish as in the English translation given here. We find no difference for the questions intended to investigate whether the subjects' perception of their tasks differed between interfaces.

Efficiency

Table 4 summarizes the time usage for the part of the essay tasks where subjects read the document, and for reading and writing the answers for the question-answering tasks.

We find a significant difference in time used for the essay tasks, $F[2,32]=4.92$, $p<.014$. A post-hoc test shows that the fisheye interface is significantly faster than the linear and the overview+detail interface; subjects complete essay tasks 16% faster.

For the question-answering tasks, we find a significant difference in time usage between interfaces, $F[2,313]=4.235$, $p<.015$. A post hoc test confirms that tasks solved with the overview+detail interface took approximately 20% longer than tasks solved with the linear interface. No difference is found between the linear and the fisheye interface.

Interface	**Essay tasks (N=58)**	**Question-answering tasks (N=354)**
Linear	44.4 (11.9) **-**	5.9 (3.5) **+**
Fisheye	37.4 (12.4) **+**	6.6 (4.3)
Overview+Detail	44.5 (12.2) **-**	7.1 (4.1) **-**

Table 4—Mean time usage in minutes for essay and each of the six questions in question-answering tasks, standard deviation is given in parenthesis. A plus denotes a significant difference to the interfaces marked with a minus at a 5% significance level.

Reading Patterns

From the logged interaction data, we are able to identify three patterns in how subjects read documents before writing essays. First, we describe subjects' reading of documents in three phases: initial orientation, linear read-through, and review (see table 5). In the initial orientation phase, subjects navigate through the document, looking especially at the abstract, the introduction, and the conclusion. In the linear read-through phase, subjects read through the document, often with regressions and skips forward to unread parts of the document. In the reviewing phase, subjects seemed to be reviewing important parts of the document. Note how only 30% of the subjects spend time in the initial orientation phase, although the fisheye interface seems to invite this behavior compared to the other two interfaces. Fewer subjects seem to be reviewing documents using the overview+detail interface and to use a smaller proportion of the total reading time to do so.

Satisfaction question	**Linear (N=20)**	**Fisheye (N=20)**	**Overview+Detail (N=19)**
Overall reaction to the system:			
Very Poor - Very Good	3.60 (1.27) -	3.68 (1.25) -	5.35 (.88) +
How was the system to use:			
Terrible - Wonderful	3.55 (1.19) -	3.74 (1.05) -	5.15 (.67) +
Hard – Easy	5.85 (1.35)	5.68 (1.29)	6.20 (.83)
Frustrating – Pleasant	3.57 (1.33) -	3.63 (1.42) -	5.55 (.83) +
Boring – Fun	3.25 (.91)	3.63 (.83)	4.57 (.94)
Confusing – Clear	5.38 (1.61)	4.58 (1.54) -	6.15 (.93) +
How do you perceived the tasks just solved:			
Very Challenging - Very Easy	4.53 (1.16)	4.79 (1.08)	4.68 (1.08)
Were your answers to the tasks:			
Very poor - Very good	4.20 (.95)	3.63 (1.12)	4.33 (.77)
How much did you learn from reading the papers:			
Learned nothing - Learning a lot	4.40 (1.23)	3.95 (1.58)	4.07 (1.13)
Were the papers just read:			
Hard to understand - Easy to understand	4.60 (1.23)	4.13 (1.33)	4.65 (1.18)
Hard to overview - Easy to overview	3.35 (1.73) -	4.05 (1.34)	5.25 (1.26) +
Was information in the two papers just read:			
Hard to locate -Easy to locate	3.95 (1.47)	4.18 (1.24)	4.65 (1.38)

Table 3—Mean scores for the 12 satisfaction questions for each interface. The first column in the table shows the question asked to the subjects (in italics), and the two extreme values showed on the seven-point differential scale that the subjects marked their answer on. Low scores were given to the negative concept of the differential scale. The next three columns show the mean scores for the three interfaces, with standard deviation given in parenthesis. A plus denotes a significant difference to the interfaces marked with a minus, using a Bonferroni adjustment of .0013.

Interface	Initial orientation	Linear read-through	Review
Linear (N=20)	4 (7 min)	20 (37 min)	13 (10 min)
Fisheye (N=19)	9 (11 min)	19 (26 min)	16 (7 min)
Overview+Detail (N=19)	4 (7 min)	19 (39 min)	10 (8 min)

Table 5—Reading phases for essay tasks. The table shows the frequency of the initial orientation, the linear read-through, and the review phase for the three interfaces. In parentheses is shown the average duration of the phase for subjects where we identified the phase. We have only counted phases that last more than 1/20 of the total reading time.

Second, we find substantial individual differences in the time used and grade obtained, in how subjects read the documents, and in which input method they used. The fastest subject spent on average 24 minutes to read the three documents used for essay tasks; the slowest subject used 2.5 times more. Incidentally, both subjects' essays received an average grade of 1.67. Two subjects read all their documents from one end to the other; four subjects used only a brief review; four subjects had both an initial orientation phase and a review phase in all of their essay tasks; and ten subjects read the documents in a more complex way. Four subjects solved all their tasks using the keyboard for input, and three subjects used only the mouse.

Third, the preferred mode of interaction for the three interfaces differs. For essay tasks, 11 subjects used mainly the arrow keys and page up/down to navigate through the document in the linear interface; three subjects used mainly the scrollbars. In the fisheye interface, subjects equally used the scrollbar and the keyboard to navigate in the document. In the overview+detail interface users are equally likely to use the scrollbar and the keyboard. However, 25% of the times subjects scroll through a document they used the overview pane as a scrollbar. While this difference superficially seems to be a natural choice of input method given the need to expand fisheye sections and the availability of a clickable overview pane, we think it might suggest differences in the way documents are read. The keyboard only allows linear navigation, while the scrollbar also allows jumping around the document.

DISCUSSION

The overview+detail interface supports reading electronic documents better than the linear and fisheye interface. The subjects' answers to essay tasks are graded higher when the overview+detail interface is used. Subjects also strongly prefer the overview+detail interface to the two other interfaces, pointing out that it supports navigation and helps to gain an overview of the structure of the document. The overview pane seems to support these activities, which pose well-known problems to readers of linear presentations of documents [22]. We think our data should encourage designers of electronic documents to use overview+detail interfaces to improve reading effectiveness and users' satisfaction.

It is puzzling that subjects use significantly more time for the question-answering tasks in the overview+detail interface compared with the other interfaces. It has been suggested that overviews impede performance for certain tasks [10,30]. We speculate that the overview pane in some situations attracts the subjects' attention, either distracting them or supporting useful associations. For the question-answering tasks, the overview pane might primarily be distracting, causing subjects to further explore the document, even when they have already found a reasonable answer to the question.

In the fisheye interface, subjects efficiently read documents for writing essays. Subjects spend less time in the linear read-through phase compared to the other interfaces. The fisheye interface seems to support subjects in efficiently grasping the main ideas using an overview oriented reading style. The subjects' satisfaction with the fisheye interface suggests that they in general do not like to depend on an algorithm that determines which sections to distort. The relatively low score for the essay tasks and the low incidental learning scores indicate that designers should be cautious in using fisheye interfaces for tasks that require a document to be fully understood. We interpret these findings to suggest that the fisheye interface is mostly useful for tasks that are time critical, for example relevance judgments.

Our study has at least five limitations, which could make the topic of further research to support reading of electronic documents with information visualization techniques. We have only considered two types of motivations for reading documents (reading-to-learn-to-do and reading-to-do); reading to judge the relevance of a document is another important activity that would be useful to support. Second, we need to consider how reading document types different from scientific documents might be supported. Third, our exploration of how reading of electronic documents might be supported should be replicated and extended for real-life reading tasks. Fourth, we think further exploration of effective semantic zooming for electronic documents is an important area for further research. While our results suggest that subjects like to be able to read the headlines of sections on the overview pane and to recognize figures and tables, it is not clear if subjects benefit from the large areas of non-readable text on the overview. Finally, we want to examine closer the individual differences in preferred reading and interaction patterns.

CONCLUSION

In an experiment, we compared the usability of three interfaces for electronic documents based on information visualization techniques. We also investigated the reading patterns of 20 subjects using these interfaces. We find that subjects prefer the overview+detail interface and with this interface write essays that receive a higher grade. Subjects complete essays faster with the fisheye interface, but seem to gain a less complete understanding of the documents read. Subjects take longer time using the overview+detail interface for answering questions, suggesting that the overview might distract them or lead to unnecessary exploration of the document. We also found different reading patterns between the interfaces. The most common interface in practical use, the linear interface, is found to be inferior to the fisheye and overview+detail interfaces regarding most aspects of usability.

Since reading of electronic documents plays a crucial role in information access and use, our results suggest that these activities might be supported through a focus on reading and interaction with electronic documents. We recommend designers of electronic documents to use overview+detail interfaces for electronic documents. Fisheye interfaces will mostly be useful for time-critical tasks when gaining a more complete understanding of the document is less important. Further research should explore individual differences in reading patterns and investigate how different reading tasks might be supported.

ACKNOWLEDGEMENTS

We thank Ben Bederson, Peter Naur, Catherine Plaisant, and Ben Shneiderman for helpful comments on a draft of this paper. We also benefited from discussions of our work at the HCIL at University of Maryland and with our colleagues at DIKU Jens Arnspang, Knud Henriksen, Kristoffer Jensen, Jørgen Sand, and Jon Sporring. We acknowledge the Digital Library Initiative at University of Illinois at Urbana-Champaign for lending us the documents used in the study. Finally, we thank the CHI reviewers for constructive comments.

REFERENCES

1. Belkin, N. J. Interaction with texts: Information retrieval as information-seeking behavior, in *Information retrieval '93. Von der Modellierung zur Anwendung* (Konstanz, Sep 1993), Universitaetsverlag Konstanz, 55-66.
2. Bishop, A. P. Working Towards an Understanding of Digital Library Use, *D-Lib Magazine*, October 1995 (1995).
3. Bishop, A. P. Digital libraries and knowledge disaggregation: the use of journal article components, in *Proceedings of Digital Libraries '98* (Pittsburgh PA, Jun 1998), ACM Press, 29-39.
4. Boguraev, B., Kennedy, C., Bellamy, R., Brawer, S., & Wong, Y. Y. Dynamic presentation of document content for rapid on-line skimming, in *AAAI Spring 1998 Symposium on Intelligent Text Summarization* (Stanford CA, Mar 1998), AAAI Press, 109-118.
5. Bradow, R., Mitze, K., & Rau, L. F. Automatic Condensation of Electronic Publications by Sentence Selection, *Information Processing & Management*, *31*, 5 (1995), 675-685.
6. Byrd, D. A scrollbar-based visualization for document navigation, in *Proceedings of Digital libraries '99* (Berkeley CA, Aug 1999), ACM Press, 122-129.
7. Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. The tangled web we wove: a taskonomy of WWW use, in *Proceeding of CHI '99* (Pittsburgh PA, May 1999), ACM Press, 544-551.
8. Card, S. K., Mackinlay, J. D., & Shneiderman, B. *Readings in Information Visualization*, Morgan Kaufmann, San Francisco CA, 1999.
9. Chen, C. & Rada, R. Interacting With Hypertext:A Meta-Analysis of Experimental Studies, *Human-Computer Interaction*, *11*, 2 (1996), 125-156.
10. Dee-Lucas, D. & Larkin, J. H. Learning From Electronic Texts: Effects of Interactive Overview for Information Access, *Cognition and Instruction*, *13*, 3 (1995), 431-468.
11. Dillon, A. *Designing Usable Electronic Text,* Taylor & Francis, London, 1993.
12. Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., & Lochbaum, C. C. Formative Design-Evaluation of SuperBook, *ACM Transaction on Information Systems*, *7*, 1 (1989), 30-57.
13. Eick, S. G., Steffen, J. L., & Sumner, E. E. Seesoft-A Tool for Visualizing Line Oriented Software Statistics, *IEEE Transactions on Software Engineering*, *18*,11 (1992),957-968.
14. Frøkjær, E., Hertzum, M., & Hornbæk, K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?, in *Proceedings of CHI 2000* (The Hague Netherlands, Apr 2000), ACM Press, 345-352.
15. Furnas, G. W. Generalized fisheye views, in *Proceedings of CHI '86* (Boston MA, Apr 1986), ACM Press, 16-23.
16. Graham, J. The reader's helper: a personalized document reading environment, in *Proceeding of CHI '99* (Pittsburgh PA, May 1999), ACM Press, 481-488.
17. Holmquist, L. E. Focus+Context Visualization with Flip Zooming and the Zoom Browser, in *CHI '97 Extended Abstracts* (Atlanta GA, Mar 1997), ACM Press, 263-264.
18. Kaugars, K. Integrated multi scale text retrieval visualization, in *CHI '98 Summary* (Los Angeles CA, Apr 1998), ACM Press, 307-308.
19. Kupiec, J., Pedersen, J., & Chen, F. A trainable document summarizer, in *Proceedings of SIGIR '95* (Seattle WA, Jul 1995), ACM Press, 68-73.
20. Marchionini, G. *Information Seeking in Electronic Environments*, Cambridge University Press, Cambridge, 1995.
21. Marcus, A. *Graphic Design for Electronic Documents and User Interfaces*, ACM Press, New York NY, 1992.
22. O'Hara, K. & Sellen, A. A comparison of reading paper and on-line documents, in *Proceedings of CHI '97* (Atlanta GA, Mar 1997), ACM Press, 335-342.
23. Páez, L. B., da Silva-Fh., J. B., & Marchionini, G. Disorientation in Electronic Environments: A Study of Hypertext and Continuous Zooming Interfaces, in *Proceedings of ASIS '96* (Baltimore MD, Nov 1996), 58-66.
24. Plaisant, C., Carr, D., & Shneiderman, B. Image Browsers: Taxonomy, Guidelines, and Informal Specifications, *IEEE Software*, *12*, 2 (1995), 21-32.
25. Robertson, G. G. & Mackinlay, J. D. The document lens, in *Proceedings of UIST '93* (Atlanta GA, Nov 1993), ACM Press, 101-108.
26. Schriver, K. *Dynamics in Document Presentation*, John Wiley & Sons, New York/London, 1997.
27. Sellen, A. & Harper, R. Paper as an analytic resource for the design of new technologies, in *Proceedings of CHI '97* (Atlanta GA, Mar 1997), ACM Press, 319-326.
28. Sparck Jones, K. & Willett, P. *Readings in Information Retrieval*, Morgan Kaufman, San Francisco CA, 1997.
29. Tomney, H. & Burton, P. F. Electronic Journals: a Study of Usage and Attitudes Among Academics, *Journal of Information Science*, *24*, 6 (1998), 419-429.
30. Wright, P. Cognitive overheads and prostheses: some issues in evaluating hypertexts, in *Proceedings of the third annual ACM conference on Hypertext* (San Antonio TX, Dec 1991), ACM Press, 1-12.

Paper 4—Navigation Patterns and Usability of Overview+detail and Zoomable User Interfaces for Maps

K. Hornbæk, B. Bederson & C. Plaisant (2001), 'Navigation Patterns and Usability of Overview+detail and Zoomable User Interfaces for Maps', 30 pages, a revised version to appear in *ACM Transactions on Computer-Human Interaction*.

This paper is a revised version of Technical Report # 2001-11, HCIL, University of Maryland, ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/2001-11html/.

Navigation Patterns and Usability of Overview+Detail and Zoomable User Interfaces for Maps

Kasper Hornbæk[*],
Department of Computing,
University of Copenhagen,
Universitetsparken 1,
DK-2100 Copenhagen Ø, Denmark,
kash@diku.dk,
+45 35321452

Benjamin B. Bederson,
Department of Computer Science,
Human-Computer Interaction Laboratory,
University of Maryland,
College Park, MD 20742, USA,
bederson@cs.umd.edu,
+1 301 405 2764

Catherine Plaisant,
Human-Computer Interaction Laboratory,
University of Maryland,
College Park, MD 20742, USA,
plaisant@cs.umd.edu,
+1 301 405 2768

The literature on information visualization establishes the usability of overview+detail interfaces, but for zoomable user interfaces, results are mixed. We compare overview+detail and zoomable user interfaces to understand the navigation patterns and usability of these interfaces. The difference between these interfaces is the presence or absence of an overview of the information space. Thirty-two subjects solved navigation and browsing tasks on maps organized in one or multiple levels. Overall, users perform better with the multi-level map. We find no difference between interfaces in subjects' ability to solve tasks correctly. Eighty percent of the subjects prefer the overview+detail interface, stating that it supports navigation and helps keep track of their position on the map. However, subjects are faster with the zoomable user interface when using a multi-level map. The combination of the zoomable user interface and the multi-level map also improves subjects' recall of objects on the map. Switching between overview and detail windows was correlated with higher task completion time, suggesting that integration of overview and detail windows require mental and motor effort.

---

[*] This work was done while the first author was visiting the Human-Computer Interaction Laboratory at the University of Maryland.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces–*Evaluation/methodology, Interaction styles*; I.3.6 [**Computer Graphics**]: Methodology and Techniques: *Interaction techniques*

General Terms: Experimentation, human factors, measurement, performance

Additional Key Words and Phrases: Information visualization, zoomable user interfaces (ZUIs), overview+detail interfaces, navigation, usability, maps, overviews

1 Introduction

Information visualization [Card et al. 1999] has become a successful paradigm for human-computer interaction. Numerous interface techniques have been proposed and an increasing number of empirical studies describe the benefits and problems of information visualization, e.g. Beard & Walker [1990], Schaffer et al. [1996], Hornbæk & Frøkjær [1999], Chen & Czerwinski [2000]. Overview+detail and zoomable user interfaces have been extensively discussed in the literature on information visualization. Overview+detail interfaces [Plaisant et al. 1995] show the details of an information space together with an overview of the entire information space. Overview+detail interfaces can improve subjective satisfaction, e.g. North & Shneiderman [2000], and efficiency, e.g. Beard & Walker [1990]. Zoomable user interfaces [Perlin & Fox 1993] organize information in space and scale, and use panning and zooming as their main interaction techniques. Research prototypes of zoomable user interfaces include interfaces for storytelling [Druin et al. 1997], web browsing [Hightower et al. 1998], and browsing of images [Combs & Bederson 1999]. However, few empirical studies have investigated the usability of zoomable user interfaces and the results of those studies have been inconclusive.

In this article we empirically investigate zoomable user interfaces by comparing them to the successful overview+detail interfaces. We compare what the literature suggests is the best implementation of both kinds of interface. However, in interfaces where zooming and panning are possible, both the zoomable user interface and the overview+detail interface should offer those interaction techniques. Not including zooming and panning would make a poor overview+detail interface and a biased comparison. In this experiment the difference between overview+detail and zoomable user interface is therefore the presence or absence of an overview. With this focus, we investigate:

- How does the presence or absence of an overview affect usability;
- How does an overview influence the way users navigate information spaces; and
- How do different organizations of information spaces influence navigation patterns and usability.

With this work we aim to strengthen the empirical literature on zoomable user interfaces, thereby identifying challenges for researchers and advising designers of user interfaces.

In the next section, we review the literature on overview+detail and zoomable user interfaces. Then, we present our empirical investigation of differences in navigation patterns and usability with and without an overview. Finally, we discuss the trade-off between time and satisfaction in such interfaces and explain the interaction between differently organized information spaces and usability.

2 Related Work

This section summarizes the research questions and empirical findings about overview+detail and zoomable user interfaces. It explains the literature behind our design decisions and motives for the experiment, both described in subsequent sections.

2.1 Overview+Detail Interfaces

Overview+detail interfaces present multiple views of an information space where some views show detailed information about the information space (so-called detail windows), while other views show an overview of the information space (so-called overview windows or overviews). Examples of overview+detail interfaces include editors for program code [Eick et al. 1992], interfaces for image collections [North et al. 1995], and commercial programs such as Adobe Photoshop[1]. Overview+detail interfaces have three benefits. First, navigation is more efficient because users may navigate using the overview window rather than using the detail window [Beard & Walker 1990]. Second, the overview window aids users in keeping track of their current position in the information space [Plaisant et al. 1995]. The overview window itself might also give users task-relevant information, e.g., by enabling users to read section titles from an overview of a document [Hornbæk & Frøkjær 2001]. Third, overview+detail interfaces give users a feeling of control [Shneiderman 1998]. A drawback of overview+detail interfaces is that the spatially indirect relation between overview and detail windows might strain memory and increase the time used for visual search [Card et al. 1999, p. 307]. In addition, overview+detail interfaces require more screen space than interfaces without overviews.

Taxonomies and design guidelines for overview+detail interfaces [Beard & Walker 1990; Plaisant et al. 1995; Carr et al. 1998; Baldonado et al. 2000] present three main findings about overview+detail interfaces. First, the overview and detail windows need to be tightly coupled [Ahlberg & Shneiderman 1994], so that navigation or selection of an information objects in one window is immediately reflected in the other windows. Tight coupling of overview and detail views has been found useful in several studies, e.g. North & Shneiderman [2000]. Second, for any relation between overview and detail windows, the zoom factor is the ratio between the larger and smaller of the magnification of the two windows. For overview+detail interfaces, this factor is recommended to be below 25 [Plaisant et al. 1995] or below 30 [Shneiderman 1998]. It is unclear, however, if the sizes of the detail and overview windows influence the recommended zoom factor. Third, the size of the overview window influences how much information can be seen at the overview and how easy it is to navigate on the overview. However, a large overview window might take screen real estate from the detail window. Plaisant et al. [1995] argue that the most usable size of the overview and detail windows is task dependent. A large overview window, for example, is required for a monitoring task, while a diagnostic task might benefit from a large detail window.

A number of studies have found overview+detail interfaces to improve user satisfaction and efficiency over detail-only interfaces. Beard & Walker [1990] compared the effect of having an overview window to navigating with scrollbars. In a 280 word

[1] See http://www.adobe.com/photoshop/

ordered tree, subjects used an overview window that allowed dragging a field-of-view and one that allowed both dragging and resizing the field-of-view. For tasks where subjects tried to locate a word in the tree and tasks where they repeatedly went from one side of the tree to the other, the overview window lead to significantly faster task completion. North & Shneiderman [2000] compared 18 subjects' performance with a detail-only, an uncoordinated overview+detail, and a coordinated overview+detail interface for browsing textual population data. Compared to the detail-only interface, the coordinated interface was 30-80% faster and scored significantly higher on a satisfaction questionnaire. Hornbæk & Frøkjær [2001] compared an overview+detail interface for electronic documents with a fisheye and a detail-only interface. Essays produced with aid of the overview+detail interface were scored significantly higher than essays produced with the aid of the detail-only interface. However, for tasks that required subjects to answer a specific question, the overview+detail interface was 20% slower compared to the detail-only interface. All but one of the 21 subjects preferred the overview+detail interface.

2.2 Zoomable User Interfaces

While zoomable user interfaces have been discussed since at least 1993 [Perlin & Fox 1993], no definition of zoomable user interface has been generally agreed upon. In this article, we consider the two main characteristics of zoomable user interfaces to be (a) that information objects are organized in space and scale, and (b) that users interact directly with the information space, mainly through panning and zooming. In zoomable user interfaces, space and scale are the fundamental means of organizing information [Perlin & Fox 1993; Furnas & Bederson 1995]. The appearances of information objects are based on the scale at which they are shown. Most common is geometric zoom, where the scale linearly determines the apparent size of the object. Objects may also have a more complex relation between appearance and scale, as in so-called semantic zooming [Perlin & Fox 1993; Frank & Timpf 1994], which is supported in the zoomable user interface toolkit Jazz [Bederson et al. 2000]. One example of semantic zooming is commonly used on maps, where the same area on the map might be shown with different features and amount of detail depending on the scale. Constant density zooming [Woodruff et al. 1998a] introduces a more complex relation between scale and appearance where the number of objects currently shown controls the appearance of objects, so that only a constant number of objects is visible simultaneously.

The second main characteristic of zoomable user interfaces is that the information space is directly visible and manipulable through panning and zooming. Panning changes the area of the information space that is visible, zooming changes the scale at which the information space is viewed. Usually, panning and zooming are controlled with the mouse or the keyboard, so that a change in the input device is linearly related to how much is panned or zoomed. Non-linear panning and zooming have been proposed in three forms: (a) goal directed zoom, where direct zooming to an appropriate scale is supported [Woodruff et al. 1998b]; (b) combined zooming and panning, where extensive panning automatically leads to zooming [Igarishi & Hinckley 2000]; and (c) automatic zoom to objects, where a click with the mouse on a object automatically zooms to center on that object [Furnas & Zhang 1998; Ware 2000]. Two ways of changing scale in a zoom action are commonly used. In jump zooming the change in scale occurs instantly, without a smooth transition. Jump zooming is used in Pad [Perlin & Fox 1993], Schaffer et al.'s [1996] experimental system, and in commercial systems such as Adobe PhotoShop or MapQuest[2]. In animated zooming the transition from the old to the new scale is smooth [Bederson & Hollan 1994; Pook et al. 2000; Bederson et al. 2000]. An important issue in animated zooming is the duration and user control over the zooming speed, i.e. the ratio between the zooming time and the zooming factor. Guo et al. [2000] provide preliminary evidence that a zoom speed around 8 factors per second is optimal. Card et al. [1991] argues that the zoom time should be approximately one second; though in some zoomable user interfaces, e.g. Jazz, users can control both the zoom time and the zoom factor. Bederson & Boltman [1999] investigated whether an animated or jump zoom technique affected 20 subjects' ability to remember the topology of and answer questions about a nine-item family tree. Subjects were better at reconstructing the topology of the tree using animated zooming, but no difference in satisfaction or task completion time was found.

The empirical investigations of zoomable user interfaces are few and inconclusive. Páez et al. [1996] compared a zoomable user interface based on Pad++ [Bederson & Hollan 1994] to a hypertext interface. Both interfaces gave access to a nine-page scientific paper. In the zoomable user interface, the scale of the sections and subsections of the paper is manipulated, so that the entire paper fits on the initial screen. No significant difference was found between the two interfaces for the 36 subjects' satisfaction, memory for the text, or task completion time. Schaffer et al. [1996] compared 20 subjects' performance with a zoomable user interface and a fisheye interface. Subjects had to locate a broken link in a telephone network and reroute the network around the link. Subjects used 58% more time for completing the task in the zoomable user interface. Subjects seem to prefer the fisheye interface, although this is not clearly described in the paper. Hightower et al. [1998] present two experiments that compare the history mechanism in Netscape Navigator with a graphical history in a zoomable user interface called PadPrints. In the first experiment, 37 subjects were required to answer questions about web pages. No significant difference in task completion time was found, but subjects preferred the PadPrints interface. In the second experiment, subjects were required to return to already visited web pages. Subjects were around 40% faster using the PadPrints interface and preferred PadPrints to Netscape Navigator. Combs & Bederson [1999] compared four image browsers: two commercial 3D interfaces, one commercial 2D interface, and an image browser based on Pad++. Thirty subjects searched for images in an image database, that they had just browsed. Subjects were significantly faster using the 2D and the zoomable user interfaces, especially as the number of images in the database went from 25 to 225. The study also presents some evidence that recall of images is improved in the zoomable user interface, but found no difference in subjective satisfaction between interfaces. Ghosh & Shneiderman [1999] compared 14 subjects' use of an overview+detail and a zoomable user interface to personal histories, LifeLines [Plaisant et al. 1996]. The zoomable user interface was marginally slower than the overview+detail interface. No difference in subjective satisfaction was found.

In general, the experimental results about zoomable user interfaces are mixed, reflecting the difference in the interfaces that zoomable user interfaces are compared to, in the organization and size of the information spaces used, and in the implementation of

[2] See http://www.mapquest.com/

zooming. In addition, the characteristics of zoomable and overview+detail interfaces are increasingly blended. Zoomable user interfaces are combined with other interfaces techniques, such as transparent overviews [Pook et al. 2000]; some overview+detail interfaces are extended with animated zooming [Ghosh & Shneiderman 1999]; and some effort has been put into extending zoomable user interfaces with navigation mechanisms that supplement zooming and panning, see for example Jul & Furnas [1998]. The main difference between research in zoomable user and overview+detail interfaces is that research in zoomable user interfaces has investigated the usefulness of zooming as a way of navigating, while overview+detail research has focused on the impact of a coupled overview. As overview+detail interfaces begin to use panning and zooming as their main navigation technique and as zoomable user interfaces begin to provide overviews and other navigational aids, the central research questions become: (1) what is the difference between different techniques for controlling and executing zooming, possibly taking into account the presence of an overview and other navigational supports; and (2) what is the effect of an overview (or other navigational supports), given that the interface provide pan and zoom techniques. In the experiment presented next, we address the latter question.

3 Experiment

To understand the differences in navigation patterns and usability between interfaces with (called overview interface) and without an overview (called no-overview interface), we conducted a controlled experiment. In the experiment, subjects used an interface with and without an overview to solve ten tasks on each of two differently organized maps.

3.1 Hypotheses

In addition to the three aims mentioned in the introduction, four hypotheses guided the design of the experiment. We hypothesized that:

(1) organization of information in multiple levels leads to more accurate and faster solutions to task with the no-overview interface compared to the overview interface. Organization of information in multiple levels provides landmarks in the information space [Vinson 1999]. This organization of information gives richer cues for where to navigate, diminishing the usefulness of the information on the overview. Thus, we hypothesize that the tradeoff between complexity and benefit of an overview in the case of multi-level information spaces will favor the no-overview interface. Note that we are not concerned, per se, with the difference between different organizations of information spaces (we expect multi-level organization to be superior), but with their interaction with interface type;

(2) recall of objects on the map would be better in the no-overview interface. Zoomable user interfaces have been speculated to improve understanding of large information spaces, because of the integrated experience of the information space [Furnas & Bederson 1995]. As mentioned in section 2, one experiment found improved recall in zoomable user interfaces. While zooming is possible in both interfaces, we expect subjects to also use the overview window for navigation in the overview+detail interface. Thus, the benefit of an integrated zoom in the detail view will be clearest in the no-overview interface;

(3) subjects prefer the overview interface, because of the information contained on the overview window and the additional navigation features. This hypothesis is based on the research on overviews in combination with non-zoomable detail views, summarized in section 2;

(4) the overview interface is faster for tasks that required comparison of information objects (e.g. by jumping between them) and scanning large areas. The literature suggests that comparison and scanning tasks are particularly well supported by an overview because the overview can be used for jumping between objects to be compared and because it can help subjects to keep track of which parts of the information space that has already been explored.

3.2 Subjects

Thirty-two subjects participated in the experiment, 23 males and 9 females. Subjects were recruited at the University of Maryland and received 15 US dollars for participating in the experiment. The age of the subjects ranged from 18 to 38; the mean age was 23.4 years. Twenty-three subjects were computer science or engineering students, four had other majors, and five were research staff or loosely affiliated with the university. Thirty-one subjects used computers every day. Twenty-three of the subjects had never used zoomable user interfaces, while nine subjects had seen or used a zoomable user interface prior to participating in the experiment. We required that subjects had spent less than two weeks in the states of Washington and Montana, because the experiment used maps of those states.

3.3 Interfaces

For the experiment, we constructed an overview and a no-overview interface; both based on the zoomable user interface toolkit Jazz [Bederson et al. 2000]. When users hold down the left mouse button, zooming in begins after a delay of 400 milliseconds. Users zoom out by holding down the right mouse button. The maximum zoom factor is 20, meaning that subjects can view the map at scale 1 through scale 20. At scale 1, the initial unmagnified view of the map is shown; at scale 20 the initial view of the map is magnified 20 times. The zoom speed is eight factors per second, i.e. subjects can zoom from the initial view of the map to the maximum magnification in 2.5 seconds. Users pan by holding down the left mouse button and moving the mouse in the opposite direction of what they wish to see (i.e. the map follows the mouse). In the lower right corner of both interfaces is an icon showing the four compass points, which are referred to in some tasks. Next to this icon is a button labeled ’zoom out’, which when pressed will zoom out to the initial view of the map. This button is expected to help subjects return to the initial view of the map if they are lost.

The no-overview interface is shown in Figure 1. Subjects may only interact with this interface using the zoom and pan techniques described above.

The overview interface is shown in Figure 2. In the top-right corner of the interface, an overview window shows the entire map at one-sixteenth the size of the detail window. This choice was arbitrary, lacking design guidelines on overview sizes (see section 2.1). However, it is similar to the average size of the overviews we are familiar with. The current location of the detail window on the map is indicated in the overview window by a 70% transparent field-of-view box. The overview and detail windows are tightly coupled, so that zooming or panning in the detail window immediately updates the overview window and dragging the field-of-view box change which part of the map is shown in the detail window. The subjects can also click in the overview window outside of the field-of-view box which will center the field-of-view box on the point clicked on. The field-of-view box can be resized by dragging the resize handle in the bottom right corner of the field-of-view box. The subjects can also draw a new field-of-view box by holding down the left button and moving the mouse until the desired rectangle has been drawn. Notice that the field-of-view box always keeps the same ratio between width and height as the detail window and overview window.

3.4 Maps

The motivation for using maps for the experiment is threefold. First, interfaces for maps constitute an important area of research. Second, maps include characteristics of other, commonly used information structures, for example hierarchical information (nesting of information objects) and network information (connections between information objects). Therefore, results concerning maps may be generalized to other information structures. Third, the direct relation between representation and physical reality aids interpretation of maps compared to the often difficult interpretation of abstract information spaces [Hornbæk & Frøkjær 1999].

We created two maps based on data from the 1995 United States Census[3].The maps contain eight types of map objects: counties, cities, parks, airports, lakes, railroads, military installations, and other landmarks. Each map object, except railroads, consists of a shape and a label. A distinct color identifies each type of map object. In addition, county names are shown in bold type and city names in italic type. Because we hypothesized that different organization of the maps might influence the navigation and usability measures, we created a multi-level and a single-level map. The maps are organized by placing labels for map objects at different scales, changing the apparent size of the labels as follows (see also Figure 3):

- The multi-level map shows map objects in the state of Washington at three levels of scale: county level (scale 1, 39 labels), city level (scale 5, 261 labels), and landmark level (scale 10, 533 labels). At the county level, labels are the same size as a 10-point font when the map is zoomed out (i.e. at scale 1) and larger when the map is magnified. When labels are shown at city or landmark level, they have the size of a 10-point font when the user has magnified the map 5 or 10 times, respectively.
- On the single-level map, all 806 labels are displayed at the scale 7, i.e. similar in size to a 10-point font when the map is magnified 7 times. The single-level map shows the state of Montana. To aid visual search, county names are also shown in capital letters.

We intended the multi-level map to be similar to information spaces that present the user with rich navigational cues everywhere in the information space (such as Yahoo style hierarchies or well designed semantic zooming). The main difference between the maps is the organization in levels. The multi-level and single-level maps are similar with respect to the number of map objects (1591 vs. 1540) and the area the state occupies (50% vs. 57% of the initial screen). The information density, measured as the mean distance in pixels from any map object to the nearest map object, is also similar (7.1 vs. 7.8). Note that a within-subjects experiment requires the use of different geographical regions for the maps, otherwise it would not be possible to control for learning effects.

[3] See http://www.census.gov/geo/www/tiger/ or http://www.esri.com/data/online/tiger/.

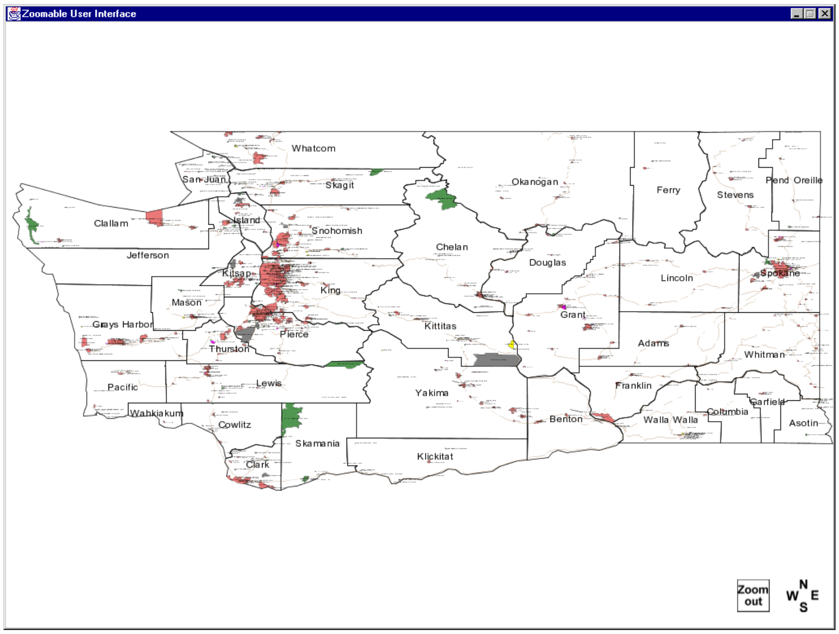


Figure 1—No-overview interface showing the multi-level map. The user may zoom and pan to change the area of the map shown. In the lower right corner of the window a button is shown that will zoom out to the initial view of the map. Next to this button is an indication of the four compass points. The colors of the map are reproduced here as different shades of gray. The map is shown at scale one, i.e. the initial view of the map.

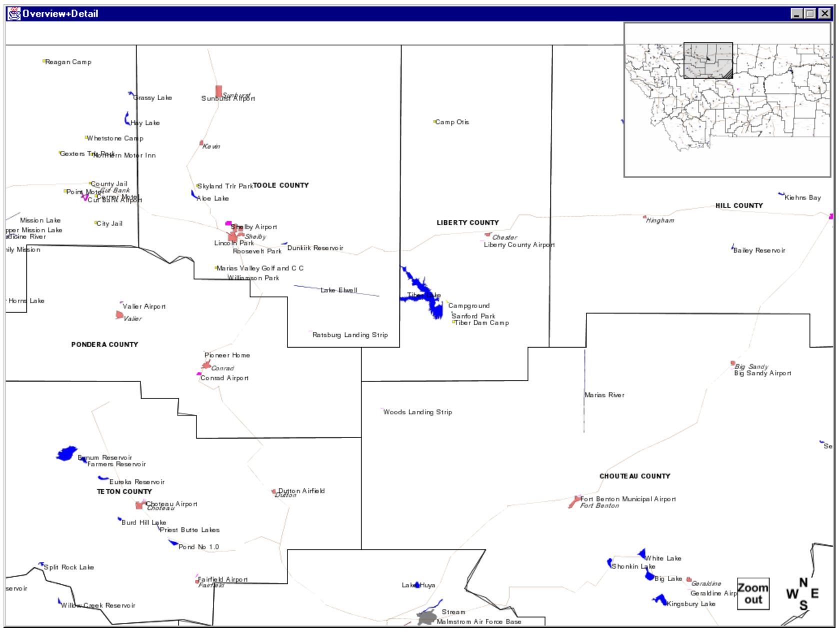
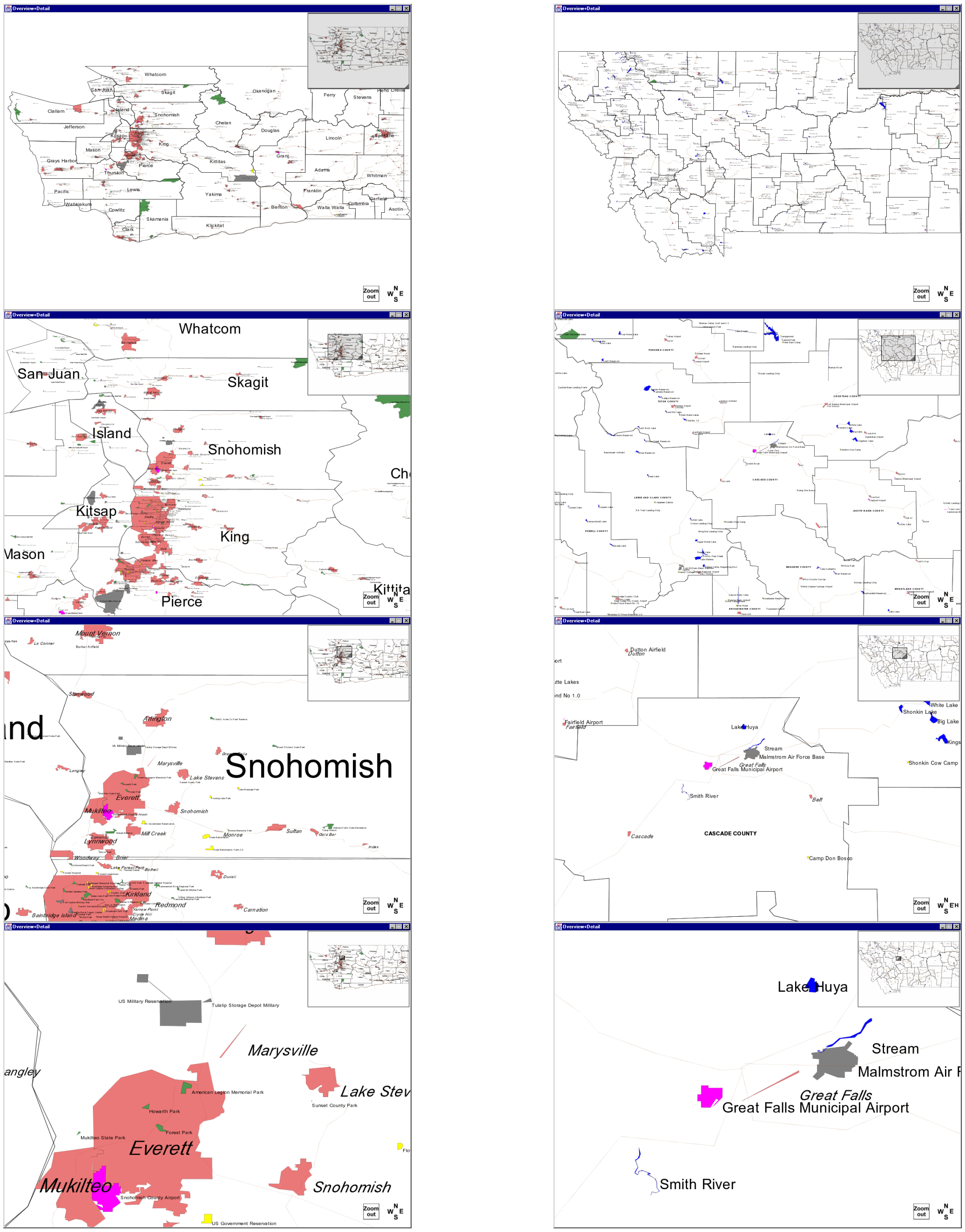


Figure 2—The overview interface showing the single-level map. In the top right corner of the interface is the overview window, which shows an overview of the entire map. The gray area in the overview window is the field-of-view box that indicates which part of the map is currently shown in the detail window. In the bottom right corner of the field-of-view box is the resize handle that allows the user to make the field-of-view smaller or larger, i.e. zooming in or out. The two buttons in the lower right corner is similar to the buttons in the zoomable user interface. The map is shown at scale four, meaning that the objects in the detail window are magnified four times.



Multi-level map **Single-level map**

Figure 3—Eight screenshots of the maps. The four screenshots in the left column show the multi-level map; the right column shows the single-scale map. From top to bottom the maps are shown at scales 1, 3, 7, and 20. On the multi-level map, map objects are labeled at three different levels: county level (39 counties, for example Snohomish in the left column, screenshot 2 from the top), city level (261 cities, for example Everett in the lower left screenshot), and landmark level (533 landmarks, barely readable in the lower left screenshot). On the single-level map, all maps objects are labeled at the same scale, i.e., all labels are same size but can appear very small at low scales. At scale 7 on this map, labels are as big as a 10-point font.

3.5 Tasks

Tasks were created to cover a large number of the types of tasks previously discussed in the literature [Plaisant et al. 1995] and to investigate specific hypotheses about when an overview would be especially useful (hypothesis 4, section 3.1). We created ten tasks for each map, five navigation tasks and five browsing tasks, which are described in the appendix.

- Navigation tasks required subjects to find a well-described map object. All of the navigation tasks specify the names of the objects to be located. In addition, the counties the objects are to be found in are named, greatly limiting the area to be searched. Two navigation tasks require subjects to locate an object on the map, two tasks require subjects to find and compare objects, and one task requires the subject to follow a route between two places specified in the task.
- Browsing tasks required subjects to scan a larger area, possibly the entire map, for objects fulfilling certain criteria. Two browsing tasks required a scan of the entire map for objects of a certain type; two tasks require subjects to scan an area of the map to find the county with most cities or the largest cities in the area, and one task required subjects to find the first object of a certain type east of some county.

Between the maps, the tasks differed only in the map objects referred to. The answers to the tasks were evenly distributed over the map, and answers were also located at different scales.

We also gave the subjects two recall tasks that test their memory of the structure and content of the map. The first recall task consisted of five small maps showing the outline of the state depicted on the map. For three of these small maps, a part of the map was darkened and the subjects were asked to write down as many objects within the dark area as they remembered. For two of the maps, subjects themselves could mark a county on the map with a cross, and write down any map objects they remembered within that county. The second recall task consisted of three county names, each associated with a list of ten cities. Subjects were told to circle all cities within a county and cross out cities they were confident were not located in the mentioned county. The list of cities consisted of the three largest cities within the county mentioned, the three largest cities in counties just next to the county mentioned, and four cities in entirely different areas of the map.

3.6 Experimental Design and Dependent Variables

The experiment varied interface type (no-overview vs. overview), map type (multi-level vs. single-level map), and task type (navigation vs. browsing tasks) within-subjects in a balanced, factorial design. The experiment consisted of two parts. In the first part, subjects used one interface giving access to one map and performed five navigation and five browsing tasks. In the second part, subjects used the other interface in combination with the not yet explored map. Subjects were randomly assigned to one of the four possible combinations of interface and map type. Within each of these four combinations, subjects were further randomly assigned to one of four permutations of task types in the two parts. Each of the resulting 16 groups contained two subjects. The order of the five tasks within a task type was the same for all subjects.

We used a range of dependent variables to capture information about usability and navigation:

- *Accuracy in answering questions.* Accuracy was calculated as the number of correct (all map objects given as answer to a task are correct), partially correct (one correct and one wrong map object), and wrong answers (all map objects are wrong).
- *Recall of map objects.* For the recall task that required subjects to mark counties and cities on the map, we counted as correct the number of counties and cities within one centimeter from the actual location of the county or city. For the recall task that required subjects to recognize the cities in a county after they had finished using the interface. We measured the number of correct indications, corrected with a penalty for guessing (the number of wrong guesses divided by the number of wrong answer possibilities for the question).
- *Task completion time.* Task completion time was measured as the time subjects could see the map. The time subjects used for the initial reading of the task, as well as the time used for entering answers, was not included.
- *Preference.* Preference was determined from subjects' indication of which interface they preferred using and from the reasons subjects gave for their indication.
- *Satisfaction.* Satisfaction was measured using seven questions with nine-point semantic differentials. Five of the questions were taken from the Questionnaire for User Satisfaction [Chin et al. 1988] and two questions were custom made. The wording of the questions appears in Figure 6.
- *Navigation actions.* We logged all interaction with the interfaces and measured the number of pan actions in the detail window and on the overview window (centering or dragging the field-of-view). We also measured zoom actions in the detail window and on the overview (resizing the field-of-view). An action is initiated when the mouse button signifying that action is pressed and is ended either when the button is released or when more than one second passes without any logged mouse movements. To compare these measures across interfaces, we combined them into a measure of total distance panned and the sum of scale changes, i.e., amount zoomed.

In section 4, this design is analyzed in two ways: (a) for measures related to task completion (accuracy, task completion time, and navigation measures), we use one solution to a task as an observation. The factors in the models used are interface (1 degree of freedom, df), map type (1 df), the interaction between interface and map type (1 df), experiment part (1 df), task type (1 df), the interaction between task and interface (1 df), subjects nested within interface and map type (60 df), and tasks nested within task type and map type (17 df). This model leaves 553 df for the error term in the multivariate analysis, 556 for univariate analysis; (b) for measures related to the use of one interface (subjective satisfaction and recall), we use as factors interface (1 df), map type (1 df), and the interaction between interface and map type (1 df), leaving 52 and 60 dfs for the multivariate and univariate analysis, respectively.

3.7 Procedure

The interfaces were run on a 650MHz Pentium III laptop with an ordinary mouse. The screen was 13 inches with a resolution of 1024*768.

Upon arriving to the lab, subjects filled out a questionnaire about gender, occupation and familiarity with computers. Then, subjects were introduced to the two interfaces and tried three practice tasks that lasted on average 11 minutes.

The main phase of the experiment consisted of two parts, each containing 10 tasks. For each task, subjects initially saw a window that covered the entire map. After reading a piece of paper that described the task, subjects clicked on a button to see a zoomed out view of the map. When subjects were ready to answer a task they entered their answer using a tightly coupled text field and list box containing the labels of all objects on the map. For all tasks, subjects were asked to proceed to the next task when they had searched for five minutes. After solving all tasks in the first part of the experiment, subjects received the recall task and filled out a satisfaction questionnaire about the interface just used. After a five-minute break, subjects began the second part of the experiment, which used the same procedure as the first part.

After the second part of the experiment, subjects filled out a form about which interface they preferred. On average, the experiment lasted one hour and 30 minutes.

4 Results

To control the experiment-wise error, we first analyzed data using a multivariate analysis of variance on all the performance measures related to task completion. We find a significant difference at the .05 significance level between interfaces, Wilk's Λ=.77, $F(4,553)=41.99$; between map types, Λ=.84, $F(4,553)=25.56$; and for the interaction between map type and interface, Λ=.942, $F(4,553)=8.56$. In a similar analysis of measures of subjective satisfaction and recall of map objects, significant differences are found for interface, Λ=.73, $F(9,52)=2.20$; between map types, Λ=.68, $F(9,52)=2.77$; and for the interaction between map type and interface, Λ=.68, $F(9,52)=2.69$. In the next sections we use univariate analyses of variance to investigate further the differences between accuracy of answers to tasks, recall of map objects, preference and satisfaction, and how subjects navigate.

4.1 Accuracy and Recall

Figure 4 summarizes the accuracy of the answers to the experimental tasks. Using a rank-based test, we find no difference in the accuracy between interfaces, $F(1,556)=.40$, $p>.5$. Between the two map types, a significant difference in the number of tasks correctly answered can be found, $F(1,556)=10.45$, $p<.001$. Tasks solved on the multi-level map are more often answered correctly than tasks solved on the single-level map.

Figure 5 shows the measures of recall of map objects for the two interfaces. With the overview interface, subjects do better at the recall task with the single-level map compared to the multi-level map. The no-overview interface shows the opposite pattern.

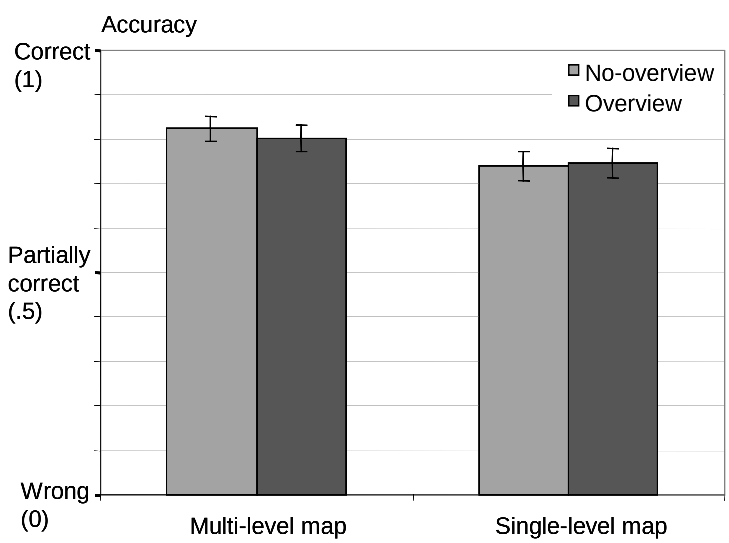


Figure 4—The average accuracy for the answers to the experimental tasks. The figure shows the average accuracy for the two interfaces between map types. The answers were scored as 1 for correct, .5 for partially correct, and 0 for wrong. A partially correct answer mentions correctly only one out of two map objects. Error bars show the standard error of the mean.

These patterns are confirmed with a rank-based test for the number of marked cities and counties by a significant interaction between interface and map type, $F(1,60)=6.96$, $p<.05$. No such interaction was found for the number of recognized cities, $F(1,60)=1.95$, $p>.1$, only a marginally significant difference between interfaces for the multi-level map was found, $F(1,60)=3.27$, $p<.08$.

Large individual differences exist in accuracy and recall of cities and counties. One subject correctly answered 19 of the 20 questions; another subject answered only nine questions correctly. In the recall task, one subject marked on average 11 cities or counties on the map; another subject marked none.

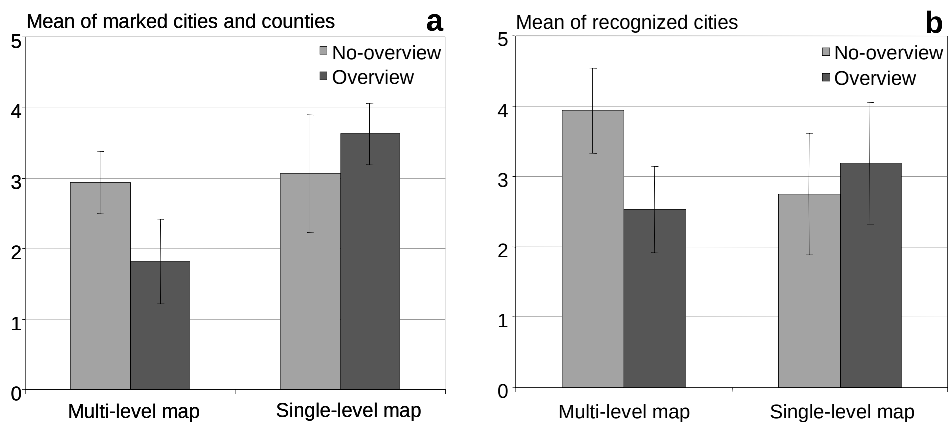


Figure 5—Mean number of correct answers to recall tasks. Panel a shows the mean number of correctly marked cities and counties; panel b shows the mean number of correctly recognized cities, adjusted for guessing. Error bars show the standard error of the mean.

4.2 Preference and Satisfaction

Twenty-six subjects stated that they preferred using the overview interface, while six subjects preferred the no-overview interface. Thus, significantly more subjects prefer the overview interface, $\chi^2(1,N=32)=12.5$, $p<.001$. Subjects explained their preference for the overview interface as follows:

- The overview window provides information about the current position on the map, for example one subject wrote “It is easier to keep track of where I am”. N=9 subjects made similar comments.
- The overview window supports navigation (N=7), for example: “[It was] easier to navigate in the overview box while looking at the detail map for answers”. Two subjects wrote similar comments at the end of the part of the experiment in which they had used the overview+detail interface.
- The overview window is helpful when scanning a large area (N=4), for example: “It made surveying a large map less disorienting especially when small landmarks had to be spotted”.
- The overview window is useful for zooming (N=2), for example “The zoom feature in the top right was extremely helpful”.

– The overview window supports comparing objects (N=2), for example: ”Easier to move between counties while at the same zoom level -> easier to compare the size of objects”.

The six subjects who preferred the no-overview interface mentioned that:

– Locating objects felt faster using the no-overview interface (N=2), for example “I found myself answering my tasks much quicker using the [no overview] interface”.
– One subject preferred the no-overview interface because the overview window got in the way when using the overview interface: “Overview+detail would seem to be more powerful, but the abundance of features got in the way to the effect of imposing on usability”. Three subjects made similar comments at the end of the part of the experiment where they used the overview+detail interface. Nevertheless, these subjects preferred the overview interface.

In addition, four subjects commented that they found it hard to resize the field-of-view box; three subjects commented that the map seemed larger using the no-overview interface; two subjects commented that when using the no-overview interface it was sometimes unclear where they were on the map; and two subjects commented that it was useful that the overview window gave a visual indication of the current zoom factor.

Figure 6 shows the subjects’ satisfaction with the overview and no-overview interfaces. The overview interface scored significantly higher than the no-overview interface on the dimensions ‘Terrible…Wonderful’, $F(1,60)=10.26$, $p<.01$; ‘Rigid…Flexible’, $F(1, 60)=7.33$, $p<.01$; and ‘Keeping track of objects were difficult…easy’, $F(1,60)=9.54$, $p<.01$). Between map types, we find a significant difference for four satisfaction questions, showing that subjects give the interfaces higher satisfaction scores when they use the multi-level map.

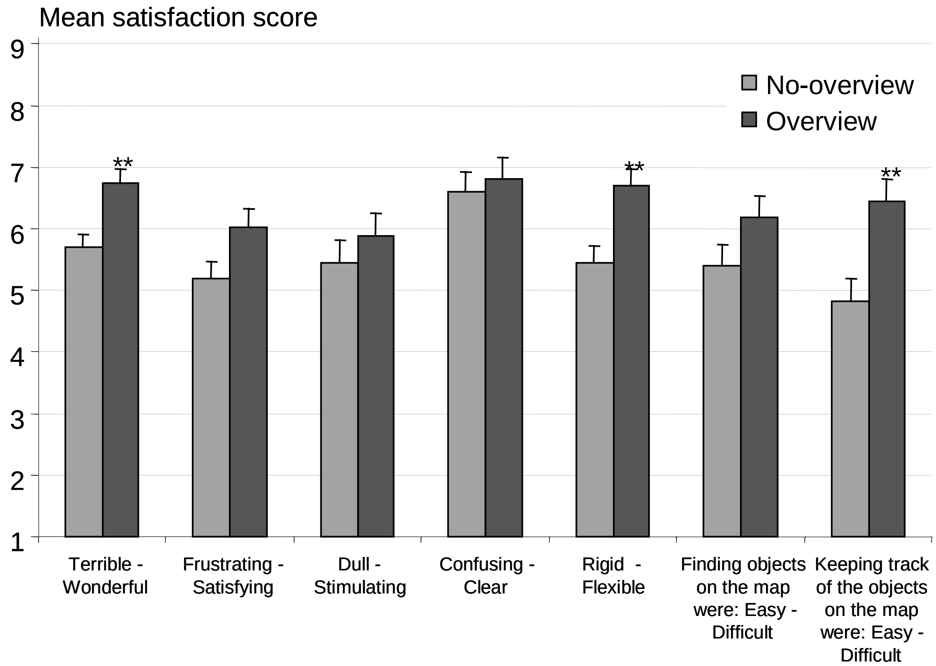


Figure 6—Satisfaction with the interfaces. The figure shows the mean score for the seven satisfaction questions in the two interfaces. Error bars indicate the standard error of the mean. The questions were answered on a nine point semantic differential going from 1 (lowest score) to 9 (highest score). Significant differences at the .01 level are marked in the figure with two asterisks (**).

4.3 Task Completion Time

Figure 7, panel **a**, shows the task completion time with the two interfaces and on the two maps. We found a significant interaction between interface and map type, $F(1,556)=6.08$, $p<.05$. Tasks solved with the no-overview interface on the multi-level map are solved 22% faster ($M=68.76$, $SD=43.38$) than tasks solved with the overview ($M=84.23$, $SD=59.42$). Tasks solved on the single-level map are solved with comparable mean completion times (No-overview: $M=107.81$, $SD=68.05$; overview: $M=105.85$, $SD=59.42$). A significant difference is also found between interfaces, $F(1,556)=4.01$, $p<.05$, indicating that the no-overview interface is faster overall. Finally, the multi-level map is faster overall compared to the single-level map, $F(1,556)=73.5$, $p<.001$.

Going into more detailed analysis, we found no significant interaction between task types and interfaces, $F(1,556)=1.98$, $p>.1$. However, as can be seen in Figure 7, panel **b**, the no-overview interface is significantly faster for navigation tasks ($M=86.9$, $SD=60.4$), compared to the overview+detail interface ($M=99.1$, $SD=64.4$), $F(1,556)=6.01$, $p<.05$.

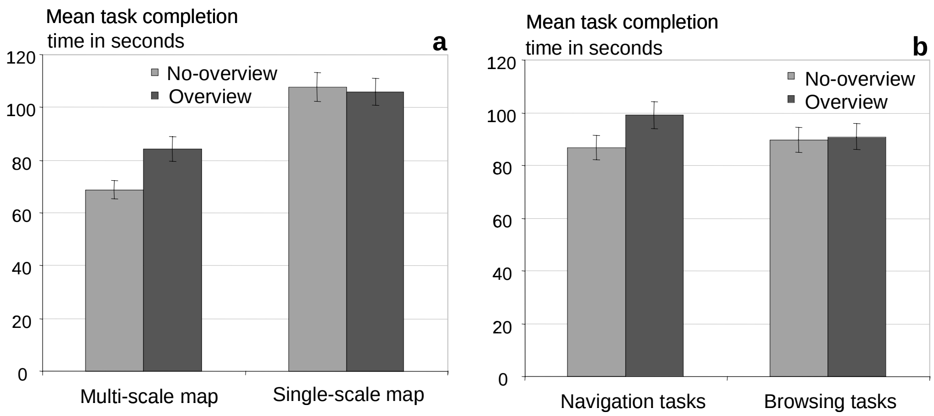


Figure 7—Task completion time in seconds. This figure shows the mean task completion time in seconds. Error bars show the standard error of the mean. Panel **a** shows the task completion time for the multi-level and the single-level map. Panel **b** shows the task completion time for navigation and browsing tasks.

All navigation tasks solved on the multi-level map with the no-overview interface had faster task completion times compared to the overview interface. Contradicting our task level hypothesis (hypothesis 4, section 3.1), we find that one of the navigation tasks that required subjects to compare map objects was solved significantly faster with the no-overview interface (estimated marginal mean=73.5, SE=11.12) compared to the overview interface (estimated marginal mean =113.9, SE=11.12), $F(1,556)=6.47$, $p<.05$. On the multi-level map, four of five browsing tasks were completed faster with the no-overview interface. One of these, a task that requires finding the first airport east of some county, is solved significantly faster using the no-overview interface (estimated marginal mean =81.81, SE=11.3) compared to the overview interface (estimated marginal mean =122.2, SE=11.2), $F(1,556)=6.20$, $p<.05$. This also contradicts our hypothesis.

For the single-level map, no significant differences between interfaces for individual tasks were found. This rejects our hypotheses that comparison tasks should be performed faster using the overview interface and that browsing tasks involving scanning the entire map should be solved faster using the overview interface.

Large differences between subjects exist. The slowest subject used on average 169 seconds per task, or 3.4 times as much as the fastest subject. For individual tasks, differences between subjects are as 1 to 23.

4.4 Navigation on the Map

In the following, we investigate the differences between navigation in the two interfaces and try to provide some data that might explain the differences in task completion time, recall tasks, and satisfaction measures discussed on the preceding pages.

4.4.1 Number of Pan and Zoom Actions

Dragging the field-of-view box is the preferred way of panning on the overview. Half of the tasks solved with the overview used this way of panning. Figure 8, panel **a**, shows the mean number of panning actions made by panning in the detail view or by centering the field-of-view. We find an interaction effect between map type and interface type, meaning that more pan actions happen with no-overview on the single-level map, F(1,556)=18.72, p<.05. However, with the overview subjects drag or center the field-of-view more frequently on the multi-level map. Hence, as can be seen in Figure 8, panel **b**, the overall distance panned, i.e. the sum of the distance panned both on the overview and on the detail view, is 51% higher with the overview (M=8690 pixels, SD=10554), compared to no-overview interface (M=5751 pixels, SD=6943), F(1,556)=22.94, p<.001.

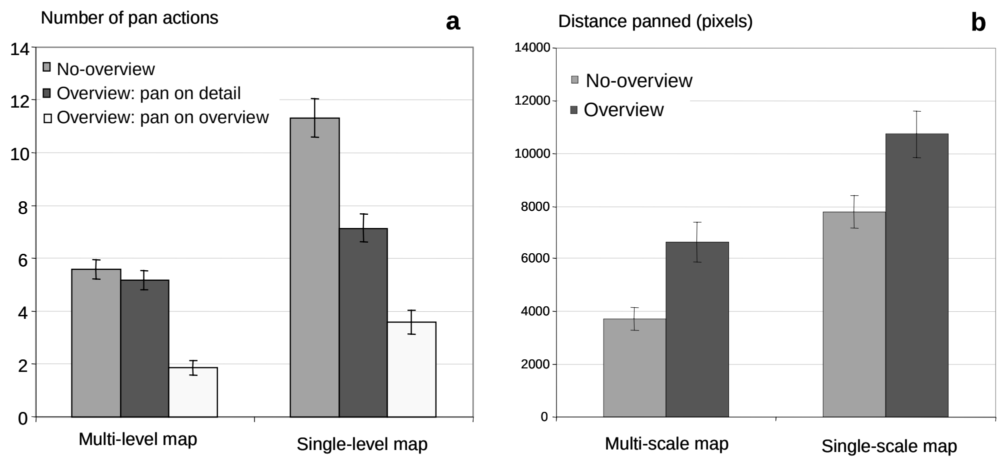


Figure 8—Panning in the two interfaces. Panel **a** shows the mean number of pan actions per task in the detail window without overview (left bar) and in the detail window with overview (middle bar), and the panning done by dragging or centering the field-of-view (right bar). Panel **b** shows the mean distance panned in screen pixels without the overview (left bar) and with the overview (right bar). In both panels, error bars show the standard error of the mean.

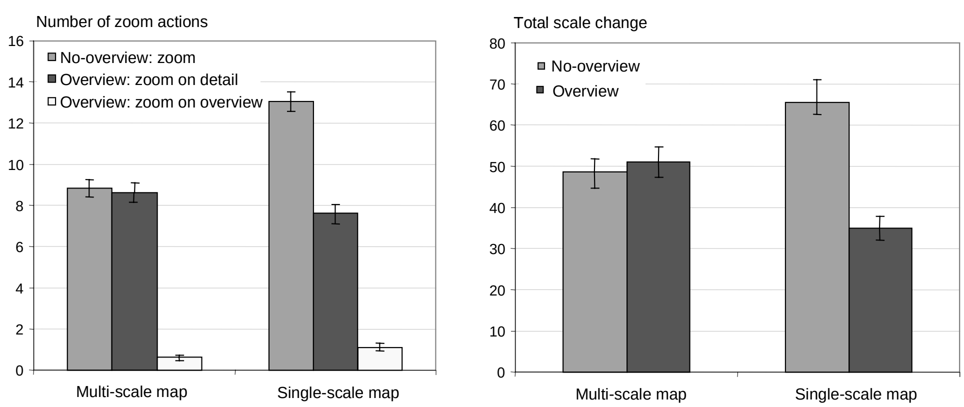


Figure 9—Zooming in the two interfaces. Panel **a** shows the mean number of zoom actions per task in the detail window without overview (left bar) and in the detail window with overview (middle bar), and the zooming done by resizing or redrawing the field-of-view (right bar). Panel **b** shows the mean scale change without the overview (left bar) and with the overview (right bar). In both panels, error bars show the standard error of the mean.

In 28% of the tasks solved with the overview, the field-of-view box is resized; in less than 4% of the tasks is the field-of-view box redrawn. Figure 9, panel **a**, summarizes the zoom actions made by resizing the field-of-view. We find a significant interaction between interface and map type, $F(1,556)=35.08$, $p<.001$, meaning that a comparable number of zoom actions is done in the two interfaces on the multi-level map, but that on the single-level map twice as much zooming happens with the no-overview interface compared to the overview interface. Subjects seldom zoom by changing the field-of-view box compared to how often they zoom on the detail view. Looking at the sum of changes in scale (Figure 9, panel **b**), we find a significant interaction between interface and map type, $F(1, 556)=25.51$, $p<.001$. On the single-level map, the no-overview interface (M=57 scales, SD=58.9) has a 33% higher number of scale changes is higher than the overview interface (M=43 scales, SD=43.2), $F(1,556)=42.79$, $p<.001$.

4.4.2 Use of the Overview Window

In 55% of the 320 tasks solved with the overview, subjects actively interacted with the overview window, i.e. they moved or resized the field-of-view box. Tasks in which the overview window was used were frequently solved by first interacting with the detail view then switching to navigating using the overview and then possibly back to the detail view. To understand better the benefit of the overview window, we compare the tasks that are solved by actively using the overview window with the tasks solved without using the overview. Tasks solved with active use of the overview are solved 20% slower (marginal mean=103.93, SE=3.98) than tasks where the overview window is not actively used (marginal mean=86.32, SE=4.57), $F(1,267)=6.75$, $p<.01$. Another way of understanding the use of the overview window is to look at the transitions between the overview and the detail window. We find that the number of transitions is strongly correlated with the time usage, Spearman's $r=.404$, $p<.001$. The more transitions between the overview and the detail window, the longer the task completion time.

Two subjects did not use the overview at all, while three subjects used the overview at least once for all ten tasks solved with the overview+detail interface.

4.4.3 Observations from the Experiment

We use our notes from observations during the experiment to make three points. First, many subjects experienced occasional problems with the combined zoom and pan button. Even though subjects practiced this combination button during the training tasks, 18 subjects at least one time zoomed when they verbally indicated that they wanted to pan. The delay before zooming begins is sometimes too short, perhaps when subjects begin initiating a pan action without having made up their minds about which direction to pan.

Second, subjects' habit formation highlighted some limitations in the interfaces. At least eight subjects tried to use a way of navigating from the overview window in the detail window or vice versa. Some subjects tried to click on the detail window, probably with the intention of jumping to the place where they clicked. This way of navigating seems to be taken from the overview window, where clicking on a point centers the field-of-view box on that point. Similarly, some subjects tried to zoom in and out while they had the mouse over the overview window. This way of interacting seems to be mimicked after the interaction with the detail view.

Third, we repeatedly observed that at least six subjects experienced what has been called desert fog [Jul & Furnas 1998], i.e. zoomed or panned into an area of the map that contained no map objects. When we observed the desert fog, two of these subjects were using the overview interface, four the no-overview interface.

5 Discussion

5.1 Usability and Navigation Patterns

Subjects preferred the overview interface. Subjects also scored this interface significantly higher on the seven satisfaction questions, and commented that the overview helped to keep track of the current position and that the overview window was useful for navigation. This result confirms our third hypotheses (see section 3.1) and is coherent with previous empirical work on overviews [North & Shneiderman 2000; Hornbæk & Frøkjær 2001] and recommendations in the design literature [Plaisant et al. 1995, Shneiderman 1998].

We found that for tasks solved on the multi-level map the interface without an overview was faster than the interface with an overview—this partially confirm our first hypothesis. We also found that subjects who actively used the overview window were slower than subjects who only used the detail window. Our results are surprising considering previous studies, e.g. Beard & Walker [1990] and North & Shneiderman [2000], which found that having an overview leads to faster task completion times. However, in the studies by Beard & Walker [1990] and North & Shneiderman [2000] navigation in the detail-only interface is done with scrollbars. Our study shows that a direct manipulation zoomable user interface and the use of a multilevel map design reduces – and possibly eliminates – the need for a separate overview. We did not find any support for our forth hypothesis about an advantage for the overview interface for certain tasks. On the contrary, when considering the difference between browsing and navigation tasks, our results are similar to those of Hornbæk & Frøkjær [2001]. In that study, as in ours, a detail-only interface was significantly faster for navigation tasks than an overview+detail interface.

In the context of our experiment, we consider four explanations of the difference in task completion time between the overview and the no-overview interfaces. First, the overview might be visually distracting, continuously catching subjects' attention and thus affecting task completion time. While we can not definitively reject this explanation from the data collected, we note that subjects who do not actively use the overview window achieved task completion times comparable to tasks solved with the zoomable user interface (see section 4.4.2). The straightforward explanation that since the interface with an overview presents more information it takes more time to use, is also weakened by this observation. A second explanation of the task completion times suggests that switching between the detail and the overview window requires mental effort and time moving the mouse. Our data modestly support this explanation, since the number of transitions between overview and detail window was positively correlated with task completion time. A third explanation is that navigation on the overview window is coarse and that resizing the field-of-view box can be difficult at low zoom factors. Subjects commented that the overview is hard to resize. In support of those comments, we note that the overview window used in the experiment occupies 256*192 pixels. When a zoom factor of 20 is reached the field-of-view box is only 13*10 pixels, which is probably hard for most users to resize and move using the mouse. Finally, it is conceivable that users never became competent in effectively using the added complexity of the overview. However, it should be noted that our experiment lasted longer than other experiments, e.g. North & Shneiderman [2000], that did found an advantage for overviews.

When using the multi-level map, subjects were faster, more accurate, and scored the interface higher on subjective satisfaction measures, irrespectively of which interface they use. The result is consistent with the literature on landmarks [Vinson 1999], since the top-level landmarks, for example the labels at the lowest scale on the multi-level map, are visible at all navigational scales. Besides being faster with the multi-level map, the no-overview interface also improved recall for map locations, partially confirming our second hypothesis. The reason for these results might be that the richer navigational cues on the multi-scale map help the subjects to concentrate navigation and attention on the detail window, thereby relying less on the overview window. Feeling lost and having to reorient oneself, possibly by using the overview window, might be less common with the multi-level map than with the single-level map.

We also set out to investigate how subjects navigated with and without an overview. Interesting, subjects only directly used the overview in half of the tasks where the overview were available. This rather low figure might indicate that adding zooming to an interface diminishes the use of the overview for navigation purposes compared to non-zoomable interfaces. Subjects panned 51% longer using the overview interface compared to the no-overview interface. One possible explanation for this large difference might be that the overview window does not support fine-grained navigation (as suggested above) and that subjects therefore have to do additional navigation on the detail view. Our data also show that subjects made more scale changes in the no-overview interface when searching the single-level map. On the single-scale map, there is less information to help navigation. The difference observed might be one indication that the overview helps both navigation and keeping an overview: a function that subjects in the no-overview condition have to substitute for more zooming.

In summary, we found a trade-off between the two interfaces, with the no-overview interface being fast and the overview interface leading to higher satisfaction. Our results challenge some of the common criticism of zoomable user interfaces without an overview, e.g. that users lose their overview when zooming [Card et al. 1999, p. 634]. We found the two interfaces to be comparable with respect to accuracy; on the multi-level map, the no-overview interface was faster than the overview interface. We do not know whether the speed difference observed might diminish when users learn to cope with the complexity of the overview interface.

5.2 Recommendations for Designers and Further Research

An interpretation of our study with the aim of providing advice for designers of information systems offers four main points. First, we found multi-level maps to be preferable to single-level maps in terms of accuracy, task completion time, and satisfaction. They should be used whenever possible.

Second, we recommend that designers closely consider the trade-off in subjective satisfaction and task completion time between providing an overview or not. We expect, in most cases, that an overview should be provided, but this depends on the critical usability parameters in the particular context designed for. A walk-up-and-use kiosk should perhaps aim for high satisfaction, while a navigation system for use in timesensitive situations could dispense with the overview if the information space contains rich cues for navigation and if the interface provides a flexible way of zooming.

Third, we believe that overview+detail interfaces should eliminate navigation commands that are specific only to the overview window or to the detail window, i.e. aim at unifying navigation [Raskin 2000]. All zoom and pan actions should therefore be similar across windows.

Fourth, to obtain the benefit of easy navigation provided by overviews (see section 2.1), designers should use overviews at least one-sixteenth the size of the detail window. For overviews coupled to a detail view less than the size of one screen or for screens on small devices, the overview might need to be larger to support navigation. For systems where much navigation is expected on the overview, for example in support of monitoring tasks, a larger overview should be provided. For systems with zoom factors over 20 as used in our system, more usability problems will occur when using the overview, and consequently a larger overview will be necessary.

We propose five areas of further research. First, the method for interacting used in the experiment occasionally causes subjects to zoom instead of pan. Experiments are needed to find a method for interacting with zoomable user interfaces using a two-dimensional input device that are intuitive and supports habit formation. We have used other interaction techniques ourselves, but picked the present interface because we believed it was easier to use for novices. Ideally, zooming and panning should be allowed to take place in parallel.

Second, empirical research should explore integrating navigational cues within the detail view. Our observations and subjects' comments suggest that a detail-only interface could include cues about the current zoom factor, e.g. Furnas et al. [2000], cues about the current position in the information space, and aids for avoiding desert fog, e.g. Jul & Furnas [1998]. If such cues are integrated into the detail view, the mental and motor effort associated with shifting to the overview might be reduced, as would the screen real estate lost due to the presence of an overview.

Third, research should aim at improving the usability of the overview window. Usability might be improved by increasing the size of the overview window or by the use of distorted overview windows, which might give users better control over local navigation without losing the possibility of coarse global navigation. Optional overviews, or space multiplexed overviews, might also provide the navigation benefit without constantly taking up screen real estate.

Fourth, in our study the use of the overview for keeping track of ones position in the information space (as opposed to using the overview for navigation) was only addressed in so far as it influenced usability. The problems users encounter when shifting visual and mental attention to the overview without interacting with it should be further explored, for example using eye tracking.

Fifth, future research could investigate in more details the effect on performance of expertise with the information space and the interface. It seems especially important to know how the satisfaction versus time tradeoff develops as users' expertise grows.

Finally, as a consequence of focusing on the effect of an overview, the second main research question about zoomable user interfaces—the difference between different techniques for executing zooming—is still largely unanswered.

6 Conclusion

We compared the navigation patterns and usability of an overview+detail interface and a zoomable user interface. The interfaces differed in whether they had an overview or not. Thirty-two subjects spent an average of one hour and 30 minutes on solving tasks on a single-level and a multi-level map. Our results suggest a tradeoff between the two interfaces in subjective satisfaction and task completion time. Subjects score the overview+detail interface higher on seven subjective satisfaction questions and 80% prefer this interface. In contrast, subjects are faster with the zoomable user interface when used with the multi-level map. Subjects prefer using the multi-level map independently of the interface used; they are also significantly faster at completing tasks on this map. We also find large individual differences in subjects' ability to navigate the map, in task completion times, and in accuracy. Based on our work, we recommend that the usability of overviews be improved, as should navigational aids for zoomable user interfaces. A better understanding of visual and mental attention in information visualization interfaces would help better explain the usability tradeoff found. Common expectations about difficulties with zoomable user interfaces and the relation between overview+detail and detail-only interfaces were not confirmed in this study. On the contrary, we found that interfaces without an overview offer certain benefits compared to interfaces with an overview.

7 Acknowledgements

We thank all members of the Human-Computer Interaction Laboratory at the University of Maryland for valuable help and encouragement. Erik Frøkjær's and Ben Shneiderman's thoughtful comments improved the paper substantially. For thorough advice on statistics, we thank Per Settergren Sørensen. Constructive and concrete advice from the anonymous reviewers is also acknowledged. This work was funded in part by DARPA's Command Post of the Future project. The work was done during a six month visit of the first author to the Human-Computer Interaction Laboratory.

Appendix: Tasks Used in the Experiment

Multi-level map, navigation tasks:

1. Which city is closest to the city Colton in Whitman County?
2. Which state park is located north of the city Ione in Pend Oreille County?
3. Which of the following two cities is located most to the north: Shelton in Mason County or Warden in Grant County?
4. Which of the following cities covers the largest area: Sequim in Clallam County, Sumas in Whatcom County, or Deer Park in Spokane County?
5. Which are the two largest parks passed on the railroad going from Westport in Grays Harbor County to Vancouver in Clark County?

Multi-level Map, browsing tasks

1. Which two national parks in Washington are biggest?
2. Find and name two counties in Washington that contain two or more military facilities.
3. Find and name the first airport east of the county Skamania.
4. Which two cities in the counties on the northern border of Washington cover the largest area?
5. Which of the counties on the southern border of Washington contains the most cities?

Single-level map, navigation tasks

1. Which city is closest to Baker City in Fallon County (in the eastern part of Montana)?
2. Which city is located west of the city Eureka in Lincoln County (in the north-west part of Montana)?
3. Which of the following two cities is located most to the north: Darby in Ravalli County (western part of Montana) or Columbus in Stillwater County (southern part of Montana)?
4. Which of the following cities in the eastern part of Montana covers the largest area: Wolfpoint in Roosevelt County, Glendive in Dawson County, or Ekalaka in Carter County?
5. Which are the two largest cities on the railroad from the city Wibaux in Wibaux County (eastern part of Montana) to the city Red Lodge in Carbon County (southern part of Montana)?

Single-level map, browsing tasks

1. Which two lakes in Montana are biggest?
2. Find and name two counties in Montana that contain at least three airports or airfields.
3. Find and name the first state park east of Furgus County (central Montana).
4. Which two cities in the counties on the northern border of Montana cover the largest area?
5. Which of the counties on the southern border of Montana contains the most cities?

8 References

Ahlberg, C. & Shneiderman, B. 1994. Visual Information Seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of ACM Conference on Human Factors in Computing Systems* (CHI '94, Boston MA, Apr. 24-28). ACM Press, New York, NY, 313-317.

Baldonado, M. Q. W., Woodruff, A., & Kuchinsky, A. 2000. Guidelines for Using Multiple Views in Information Visualization. In *Proocedings of the 5th International Working Conference on Advanced Visual Interfaces* (AVI'2000, Palermo, Italy, May 24-26). Tarrantino, L. Ed. ACM Press, New York, NY, 110-119.

Beard, D. B. and Walker, J. Q. 1990. Navigational Techniques to Improve the Display of Large Two-Dimensional Spaces. *Behaviour and Information Technology, 9,* 6, 451-466.

Bederson, B. B. & Boltman, A. 1999. Does Animation Help Users Build Mental Maps of Spatial Information. In *Proceedings of IEEE Symposium on Information Visualization* (INFOVIZ'99, San Fransisco, CA, Oct. 24-29). IEEE Press, New York, NY, 28-35.

Bederson, B. B. & Hollan, J. D. 1994. Pad++: A Zooming Graphical Interface System. In *Proceedings of the 7th ACM Symposium on User Interface Software and Technology* (UIST'94, Marina del Rey,CA, Nov. 2-4). ACM Press, New York, NY, 17-26.

Bederson, B. B., Meyer, J., & Good, L. 2000. Jazz: An Extensible Zoomable User Interface Graphics ToolKit in Java. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (UIST'00, San Diego, CA, Nov. 6-8). ACM Press, New York, NY, 171-180.

Card, S. K., Mackinlay, J. D. & Shneiderman, B. 1999. *Readings in Information Visualization.* San Francisco CA: Morgan Kaufmann.

Card, S. K., Robertson, G. G., & Mackinlay, J. D. 1991. The information visualizer, an information workspace. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI'91, New Orleans, LA, Apr. 27- May 2). ACM Press, New York, NY, 181-188.

Carr, D., Plaisant, C., and Hasegawa, H. 1998. Designing a real-time telepathology workstation to mitigate communication delays. *Interacting with Computers, 11,* 1, 33-52.

Chen, C. and Czerwinski, M. P. 2000. Special Issue on Empirical Evaluation of Information Visualizations. *International Journal of Human-Computer Studies, 53,* 5.

Chin, J. P., Diehl, V. A., & Norman, K. L. 1988. Development of an instrument for measuring user satisfaction of the human-computer interface. In *Proceeding of the ACM Conference on Human Factors in Computing Systems* (CHI '88, Washington, DC, May 15-19). Soloway, E., Frye, D., and Sheppard, S. B. Eds. ACM Press, New York, NY, 213-218.

Combs, T. & Bederson, B. B. 1999. Does Zooming Improve Image Browsing? In *Proceedings of the ACM Conference on Digital Libraries* (DL '99, Berkeley, CA, Aug. 11-14). ACM Press, New York, NY, 130-137.

Druin, A., Stewart, J., Proft, D., Bederson, B., & Hollan, J. D. 1997. KidPad: A Design Collaboration Between Children, Technologists, and Educators. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '97, Atlanta, GA, Mar. 22-27). Pemperton, S. Ed. ACM Press, New York, NY, 463-470.

Eick, S. G., Steffen, J. L., and Sumner, E. E. 1992. Seesoft-A Tool for Visualizing Line Oriented Software Statistics. *IEEE Transactions on Software Engineering, 18,* 11, 957-968.

Frank, A. U. and Timpf, S. 1994. Multiple Representations for Cartographic Objects in a Multi-Scale Tree-An Intelligent Graphical Zoom. *Computers & Graphics, 18,* 6, 823-829.

Furnas, G. W. & Bederson, B. B. 1995. Space-Scale Diagrams: Understanding Multiscale Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '95, Denver, CO, May 7-11). Katz, I. R., Mach, R., Marks, L., Rosson, M. B., and Nielsen, J. Eds. ACM Press, New York, NY, 234-241.

Furnas, G. W. & Zhang, X. 1998. MuSE: a multiscale editor. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (UIST '98, San Fransisco, CA, Nov. 1-4). ACM Press, New York, NY, 107-116.

Furnas, G. W. & Zhang, X. 2000. Illusions of infinity: feedback for infinite worlds. In *Proceedings of the 13th Annual ACM symposium on User Interface Software and Technology* (UIST 2000, San Diego, CA, Nov. 5-8). ACM Press, New York, 237-238.

Ghosh, Partha and Shneiderman, Ben 1999. Zoom-Only vs Overview-Detail pair: A study in Browsing Techniques as Applied to Patient Histories. *University of Maryland Technical Report, CS-TR-4028,* ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/99-12html/99-12.html.

Guo, Huo, Zhang, Weiwei, and Wu, Jing 2000. The Effect of Zooming Speed in a Zoomable User Interface. *Report from Student HCI Online Research Experiments (SHORE),* http://otal.umd.edu/SHORE2000/zoom/.

Hightower, R. R., Ring, L. T., Helfman, J. I., Bederson, B. B., & Hollan, J. D. 1998. Graphical Multiscale Web Histories: A Study of PadPrints. In *Proceedings of the Ninth ACM Conference on Hypertext* (Hypertext '98, Pittsburgh, PA, June 20-24). ACM Press, New York, NY, 58-65.

Hornbæk, K. & Frøkjær, E. 1999. Do Thematic Maps Improve Information Retrieval? In *IFIP TC.13 International Conference on Human-Computer Interaction* (INTERACT '99, Edingburgh, Scotland, Aug. 30- Sep. 3). IOS Press, Amsterdam, 179-186.

Hornbæk, K. & Frøkjær, E. 2001. Reading Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2001, Seattle, WA, Mar. 31- Apr. 5). ACM Press, New York, NY, 293-300.

Igarishi, T. & Hinckley, K. 2000. Speed-dependent automatic zooming for browsing large documents. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (UIST 2000, San Diego, CA, Nov. 5-8). ACM Press, New York, NY, 139-148.

Jul, S. & Furnas, G. W. 1998. Critical zones in desert fog: aids to multiscale navigation. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (UIST '98, San Fransisco, CA, Nov. 1-4). ACM Press, New York, NY, 97-106.

North, C. and Shneiderman, B. 2000. Snap-Together Visualization: Evaluating Coordination Usage and Construction. *International Journal of Human Computer Studies, 53,* 715-739.

North, C., Shneiderman, B., & Plaisant, C. 1995. User controlled overviews of an image library: A case study of the Visible Human. In *Proceedings of the 1st ACM*

International Conference on Digital Libraries (DL '96, Bethesda, MD, Mar. 20-23). ACM Press, New York, NY, 74-82.

Páez, L. B., da Silva-Fh., J. B., & Marchionini, G. 1996. Disorientation in Electronic Environments: A Study of Hypertext and Continuous Zooming Interfaces. In *Proceedings of the 59th Annual Meeting of the American Society for Information Science* (ASIS '96, Baltimore, MD, Oct. 19-24). Harding, S. Ed., 58-66.

Perlin, K. & Fox, D. 1993. Pad: An Alternative Approach to the Computer Interface. In *Proceedings of the 20th Annual ACM Conference on Computer Graphics* (SIGGRAPH '93, Anaheim, CA, Aug. 2-6). Kajiya, J. T. Ed. ACM Press, New York, NY, 57-64.

Plaisant, C. M. B., Rose, A., & Shneiderman, B. 1996. Life Lines: Visualizing personal histories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '96, Vancouver, Canada, Apr. 13-18). ACM Press, New York, NY, 221-227.

Plaisant, C., Carr, D., and Shneiderman, B. 1995. Image browsers: Taxonomy, guidelines, and informal specifications. *IEEE Software, 12,* 2, 21-32.

Pook, S., Lecolinet, E., Vaysseix, G., & Barillot, E. 2000. Context and Interaction in Zoomable User Interfaces. In *Proocedings of the 5th International Working Conference on Advanced Visual Interfaces* (AVI 2000, Palermo, Italy, May 23-26). ACM Press, New York, NY, 227-231.

Raskin, J. 2000. *The Humane Interface: New Directions for Designing Interactive Systems.* Reading, MA: Addison-Wesley.

Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., and Roseman, M. 1996. Navigating Hierarchically Clustered Networks through Fisheye and Full-Zoom Methods. *ACM Transactions on Computer-Human Interaction, 3,* 2, 162-188.

Shneiderman, B. 1998. *Designing the User Interface.* Reading, MA: Addison-Wesley.

Vinson, N. G. 1999. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '99, Pittsburg, PA, May 15-20). Williams, M. G., Altom, M. W., Ehrlich, K., and Newman, W. Eds. ACM Press, New York, NY, 278-285.

Ware, C. 2000. *Information Visualization: Perception for Design.* San Fransisco, CA: Morgan Kaufmann Publishers.

Woodruff, A., Landay, J., & Stonebreaker, M. 1998. Constant Information Density in Zoomable Interfaces. In *Proocedings of the 4th International Working Conference on Advanced Visual Interfaces* (AVI '98, L'Aquila, Italy, Maya 24-27). 110-119.

Woodruff, A., Landay, J., & Stonebreaker, M. 1998. Goal-Directed Zoom. In *Summary of the ACM Conference on Human Factors in Computing Systems* (CHI '98, Los Angeles, CA, Apr. 18-23). Karat, C.-M., Lund, A., Coutaz, J., and Karat, J. Eds. ACM Press, New York, NY, 305-306.

Paper 5—Reading Patterns and Usability in Visualizations of Electronic Documents

K. Hornbæk & E. Frøkjær (2001), 'Reading Patterns and Usability in Visualizations of Electronic Documents', 26 pages, under review.

Reading Patterns and Usability in Visualizations of Electronic Documents

Kasper Hornbæk,
Department of Computing,
University of Copenhagen,
Universitetsparken 1,
DK-2100 Copenhagen, Denmark,
Phone +(45) 35321452,
Fax +(45) 35321401,
E-mail kash@diku.dk

Erik Frøkjær,
Department of Computing,
University of Copenhagen,
Universitetsparken 1,
DK-2100 Copenhagen, Denmark,
Phone +(45) 35321456,
Fax +(45) 35321401,
E-mail erikf@diku.dk

Abstract:
We present an exploration of reading patterns and usability in visualizations of electronic documents. Twenty subjects wrote essays and answered questions about scientific documents using an overview+detail, a fisheye, and a linear interface. We study reading patterns by progression maps that visualize the progression of subjects' reading activity; and visibility maps that show for how long different parts of the document are visible. The reading patterns help explain differences in usability between the interfaces and show how interfaces affect the way subjects read. With the overview+detail interface, subjects get higher grades for their essays. All but one of the subjects prefer this interface. With the fisheye interface, subjects use more time on gaining an overview of the document and less time on reading the details. Thus they read the documents faster, but display lower incidental learning. We also show how subjects only briefly have visible the parts of the document that are not initially readable in the fisheye interface. This happens even though subjects express a lack of trust in the algorithm underlying the fisheye interface. When answering questions, the overview is used for jumping directly to answers in the document and to already-visited parts of the document. However, subjects are slower at answering questions with the overview+detail interface. From the visualizations of the reading activity, we find that subjects using the overview+detail interface often explore the document further even when a satisfactory answer to the given question has already been read. Thus overviews occasionally grab subjects' attention and possibly distract them.

Keywords:
Electronic documents, digital documents, information retrieval, information visualization, reading, reading patterns, overview+detail interface, fisheye interface

1 Introduction

Reading of electronic documents has become ubiquitous and deeply integrated in our everyday activities. Such documents are read on the World Wide Web, in electronic journals, in professional work, and as part of recreational activities. Sellen & Harper [1997] describe the use of paper and electronic documents among analysts at the International Monetary Fund and assess that 14% of the time analysts worked with documents they used electronic documents only. Analysts used a combination of paper and electronic documents 35% of the time. Byrne et al. [1999] studied World Wide Web usage and found that users spend at least twice as much time using the information they find, compared to searching, browsing, or any other activity. In the study of Byrne et al., reading is the main activity in using information.

Unfortunately, users experience a variety of difficulties when reading electronic documents. These difficulties include cumbersome navigation [Dillon 1994; O'Hara & Sellen 1997], a lack of overview of the document [O'Hara & Sellen 1997], lower tangibility of electronic documents compared to paper [Hansen & Haas 1988], an unclear awareness of the length of documents [O'Hara & Sellen 1997], lower reading speed caused by the poor resolution of most screens [Mills & Weldon 1987; Dillon 1994], learning of lower quality compared to paper documents [Hertzum & Frøkjær 1996], and possible fatigue if reading for extended periods of time.

As a potential solution to these problems and with the aim of improving the ubiquitous reading activity, visualization techniques have been used for presenting electronic documents [Eick et al. 1992; Hornbæk & Frøkjær 2001]. Some visualizations of electronic documents show the contents of a document together with an overview of that document [Eick et al. 1992; Graham 1999]. Others show a distorted version of the document compressed to fit a limited amount of screen space [Robertson & Mackinlay 1993] or consisting of only the important parts of the document [Furnas 1986; Kaugars 1998]. However, the usability of visualizations of electronic documents is largely unexamined and to our knowledge no one has investigated if such interfaces change how users read.

In this paper, we analyze how visualization techniques support reading of electronic documents. We compare a linear, a fisheye, and an overview+detail interface used in an experiment by 20 subjects for writing essays and answering questions about scientific documents. We use logged data about the interaction process to visualize subjects' reading activity. Our visualizations help describe reading patterns by showing how reading progresses and for how long certain parts of a document are visible. The reading patterns give insight into how the interfaces affect subjects' reading activity and into how we can design interfaces that better support reading. In addition, we investigate the common hypothesis that overview+detail and fisheye interfaces improve usability. Extending our previous analysis [Hornbæk & Frøkjær 2001], we use the reading patterns to explain differences in usability between the interfaces. This gives rise to some hypotheses about how the visualizations affect subjects mentally.

In the next section, we outline previous work on visualization and on studies of reading patterns in electronic documents. Section 3 describes our experiment on visualizing electronic documents. Section 4 describes the reading patterns. Section 5 present the differences in usability between interfaces and explain them with reference to the reading patterns. In section 6 the results are discussed and section 7 presents our main conclusions.

2 Background

2.1 Visualization of Electronic Documents

Visualizations of electronic documents are of two kinds: overview+detail and distortion-based interfaces. Overview+detail interfaces show an overview of the document separated from the detailed content [Plaisant et al. 1995]. The overviews show zoomed out representations of the document [Eick et al. 1992; Boguraev et al. 1998; Graham 1999] or thumbnail representations of the pages in the document [Adobe Acrobat[1]; Ginsburg et al. 1996]. On some overviews occurrences of query terms in the document are colour coded [Graham 1999; Byrd 1999]. Besides the present paper, we know of no evaluations of overview+detail interfaces for electronic documents. However, Chen & Rada [1996]'s review of research in hypertext suggests that overviews improve the users' effectiveness. Studies of text overviews also suggest improved performance from having an overview of an electronic document. Three studies of Superbook [Egan et al. 1989] compared the performance of subjects who used a 562-pages paper manual for a statistics package to subjects searching an electronic version of the manual using an expandable table of contents (i.e. a text overview) combined with the detailed contents of the manuals. In the third study, 10 subjects performed 25% better with Superbook than subjects searching in paper manual. In two experiments Dee-Lucas & Larkin [1995] compared linear text to overview interfaces in which the overview and the detailed contents were not visible simultaneously. When reading an approximately 2000-words physics text, the subjects using the overview had better and broader recall of text topics compared to subjects without the overview.

Distortion-based interfaces show the entire document in a limited amount of screen space or show only the most important parts of the document. Robertson & Mackinlay [1993] proposed an interface that shows only one part of a document in focus and the other pages of the document zoomed out to fit the remaining space. Holmquist [1997] describes a similar interface that can use semantic zooming on the pages that are out of focus. In other distortion-based interfaces only important parts of the document are readable. Importance may be determined by structural properties of the document, such as sections and subsections [Páez et al. 1996]; by the current view of the document [Furnas 1986]; or by similarity between the terms used for retrieving the document and the sections of the document [Kaugars 1998]. Páez et al. [1996] describe a zoomable user interface for electronic documents where title, headings, and key sentences are larger than other parts of the document. Initially, the entire document is visible on the screen. When comparing this interface to a hypertext interface, Páez et al. [1996] found no difference between interfaces in 36 subjects' satisfaction, task completion time, or memory for the document.

2.2 Reading Patterns

A large literature describes how interface designs, tasks, genre characteristics, and reader traits influence performance when reading electronic documents [Wright 1987; Hansen & Haas 1988; Dillon 1994; Muter 1996; Schriver 1997]. Here we focus on characterizing patterns in reading activity, i.e. how readers navigate and manipulate documents as they try to accomplish their aims with reading. Three kinds of reading patterns are discussed in the literature.

As one reading pattern, documents are read in a non-linear fashion, occasionally with multiple readings of some sections. Bazerman [1988, p. 235-253] discusses how the purposes and background knowledge influence the way seven physicists read academic papers. In general, papers were read selectively with jumps between different sections. Readers often looked for new

[1] http://www.adobe.com/products/acrobat

information or for particular sections, such as the method section in descriptions of empirical research. In addition, parts of the documents were given multiple readings at different intensity. Dillon [1994, p. 93-101] describes two series of 15 interviews about how participants read academic papers and software manuals. For academic papers, most readers skim titles and author names, after which they scan the abstract and main sections. Then, important sections are read non-linearly or the whole paper is read serially. In software manuals, the participants most often consulted the table of contents or the index sections to get a feel for the contents and locate useful places for reading. Horney & Anderson-Inman [1994] describe the reading patterns of 17 middle school students in two hypertext stories. From logged interaction with the stories, they identify different processes in the reading activity such as skimming, checking, reading, responding, studying, and reviewing. Horney & Anderson-Inman [1994] also show how students read the stories multiple times and how students sometimes read the story from end to beginning.

As a second reading pattern, linear reading occurs under some circumstances. Goldman & Saul [1990] showed that the most common reading strategy among students reading informational texts was to read linearly through the text once. Foltz [1996] compared the reading strategies in two hypertexts and a linear document. Independently of task type (reading for general knowledge vs. reading for finding specific information) and document type, 80 to 90% of the transitions to new sections and pages were coherent with the overall organization of the text. In a second experiment, Foltz used verbal reports to show that when subjects answered specific questions they read linearly from text preceding the desired information and towards that information, apparently trying to maintain the coherence of the text. Similarly, subjects in the experiment of Hertzum et al. [2001] often begin reading sections preceding the section containing the answer to the question posed. Seemingly, subjects try to establish the context of the answer.

A third group of reading patterns is formed by the various roles played in reading by different parts of a document. For academic papers, certain sections, e.g. those containing dense formulas or problem formulations, might be skipped entirely [Bazerman 1988]. Bishop [1999] used focus groups and interviews to investigate how readers of scientific papers use document components. She shows how readers use document components, such as the abstract or figures, for orientation, for gaining an overview of the paper, for directing attention, for comprehension, and for inspiring additional reading. In addition, readers often jump non-linearly between different parts of the paper.

In summary, reading patterns are diverse and no one has studied reading patterns for visualizations of electronic documents. New in this study are therefore the investigations of reading patterns in overview+detail or distorted interfaces and the detail of the descriptions of reading patterns.

3 Experiment

To investigate how visualizations of electronic documents influence reading patterns, we conducted an experiment where subjects answered questions and wrote essays about documents on object-oriented systems development. Subjects completed these tasks using a linear, a fisheye, and an overview+detail interface. Below we describe the interfaces and the experiment; Hornbæk & Frøkjær [2001] contained a preliminary account of the usability data from the experiment but only a brief mention of reading patterns, our main focus here.

Our experiment is exploratory, aiming at describing reading patterns and how interfaces affect reading. In addition to this aim, we had two hypotheses about differences between interfaces.

1. Based on the literature described in section 2 we expected the overview+detail interface to improve satisfaction and task completion time over the linear interface. We expected this because the overview+detail interface facilitates navigation by providing the overview pane and because this interface presents the reader with an overview of the structure and contents of the entire document.
2. We also expected the fisheye interface to decrease task completion time because the documents are compressed in the presentation and therefore less time-consuming to navigate. The fisheye interface was also expected to support readers in employing an overview-oriented reading style, so-called outlining [Anderson & Armbruster 1982]. One measurable implication of this is faster reading, since subjects quickly establish an overview of the text.

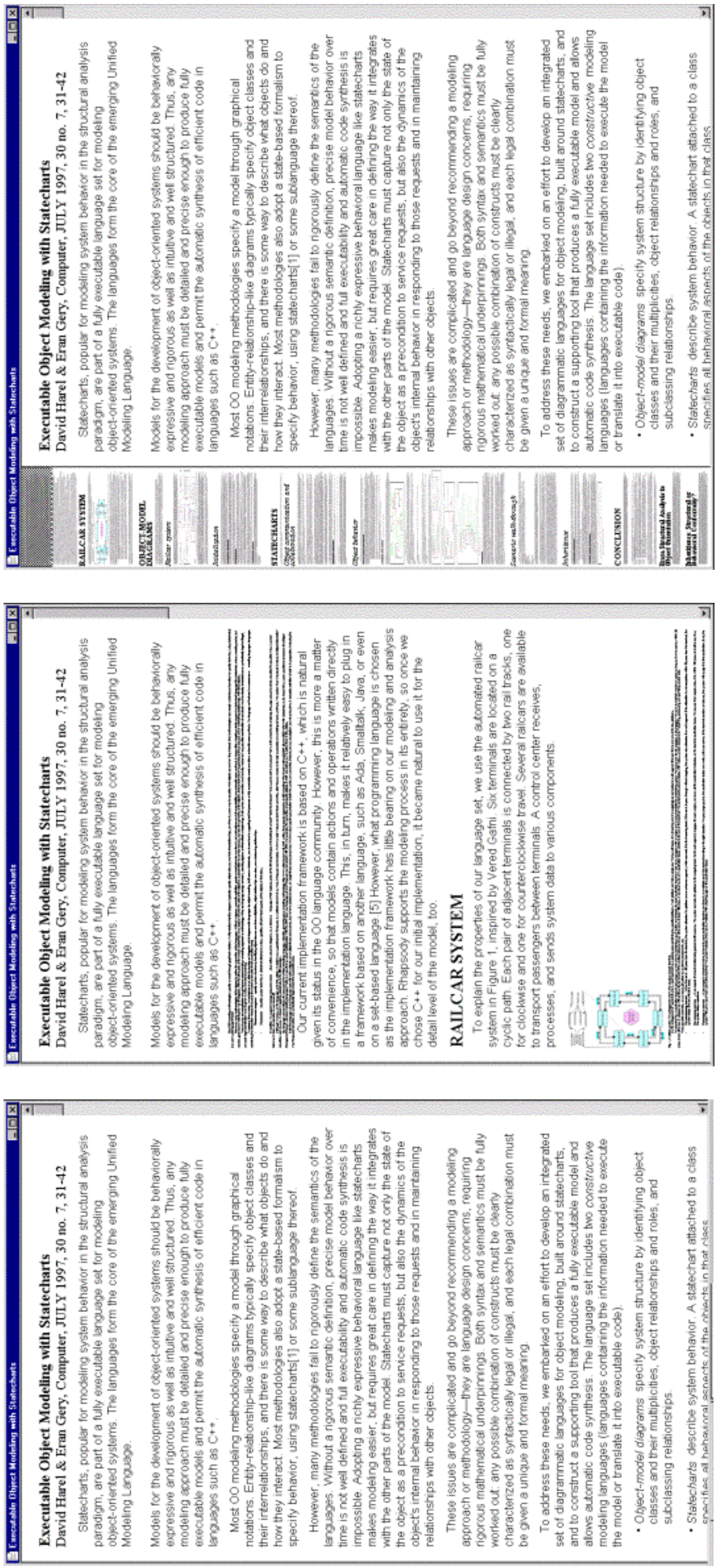
3.1 Interfaces

We compared a linear, a fisheye, and an overview+detail interface. In these interfaces, documents can be navigated using the mouse or the keyboard. Subjects may highlight words in the documents. By entering one or more words in a dialog box, all instances of the entered words are highlighted in red in the document. Figur 1 shows the three interfaces.

In the linear interface, the document is shown as a linear sequence of text and pictures. This interface is similar to most interfaces in practical use and serves as a baseline against which the other interfaces can be compared.

In the fisheye interface, certain parts of the document are considered more important than other parts. The most important parts of a document are always readable. The other parts of the document are initially distorted below readable size, but can be expanded and made readable if the user clicks on them with the mouse. Because of the distortion, the initial size of the documents in the fisheye interface was on average 25% of their sizes in the linear interface. Two strategies are used for determining which sections are important. First, sentences selected from the beginning and end of a document unit are among the best indicators of the contents of that unit [Bradow et al. 1995; Kupiec et al. 1995]. Therefore, the first and last paragraphs of a section are considered important. This scheme is recursively applied to subsections, so that when a section is expanded only the first and last parts of the subsections are readable. Second, as mentioned in section 2 readers often attend to and find certain components of a document especially useful [Dillon 1994; Bishop 1999]. Therefore abstracts and section headings are always visible, and graphics and tables are diminished less than text.

In the overview+detail interface, the document is shown as a linear sequence of text and pictures (the detail pane) together with a tightly coupled overview of the document (the overview pane). For the six documents used in the experiment, the ratio between the length of the overview pane and the length of the entire document was on average 1:17. A rectangular field-of-view covering a part of the overview pane indicates which part of the document is currently shown in



Figur 1—The interfaces. This figure shows from bottom to top the linear, the fisheye, and the overview+detail interface.

the detail pane. The field-of-view can be moved to change which part of the document is shown in the detail pane. On the overview pane, section and subsection headings are shown at a fixed size. Except for the headings, the contents of a section are zoomed to fit the remaining space allocated to show that section. We believe that the readability of headings and the stability of the overview pane are the main improvement over previous overview+detail interfaces for electronic documents, e.g. Graham [1999].

3.2 Tasks and Documents

Subjects were given two types of tasks: essay tasks and question-answering tasks. The essay tasks and the question-answering tasks correspond to reading to understand a document and reading to answer a question. These aims of reading are central in several accounts of typical reading tasks, e.g. Schriver [1997]. Although answering questions is obviously a typical task with electronic documents, it may be argued that no one reads an entire document from the screen. However, our intention with the overview+detail and the fisheye interface is to make online reading more attractive and thus we need to look at tasks that make subjects read to understand.

In essay tasks, subjects read a document to learn the main contents of that document. Afterwards and without access to the document, they were required to write a one-page essay, stating the main theses and ideas of the document, and one page of personal comments about the document. After writing the essays, subjects were given six incidental-learning questions. An example of an incidental-learning question is: ‘Which integrity problems can occur in what the author calls the simple business application architecture?’

In question-answering tasks, subjects were required to answer six questions about a document, one question at a time. The six questions were varied as to (1) position in the document where the answer can be found (in the first or last part of the document), (2) how easily accessible the sentences or sections containing the answer are (whether they are near section beginnings, tables or figures), and (3) the usefulness of the words of the question as terms for highlighting (whether or not the question contained terms that were located near the answer). An example of a question is: ‘What is, according to the paper, the biggest problem in relation to automatically transforming procedural code to object-oriented code?’

The documents used in the experiment were six IEEE journal papers from the Digital Library Initiative test bed at University of Illinois at Urbana-Champaign [Bishop 1995]. All documents were on topics within object-oriented systems development. The paper versions of the documents were between 8 and 14 pages long. The documents contained figures, tables, formulas, and text. From our presence during the experiment we conclude that no subjects had previously read any of the papers. For uninterrupted reading and increased realism, we did not impose a time limit on the tasks. However, subjects were made aware of how much time they had used when reading one paper for more than one hour, or when they took more than 30 minutes to answer one of the six questions about a document. The descriptions of the tasks, the answers to the tasks, the training material, and the satisfaction questionnaires were all in the native language of the subjects, Danish.

3.3 Subjects

The subjects in the experiment were students at the Department of Computing, University of Copenhagen, who chose to participate in a course involving the experiment. The subjects had studied computer science for a mean time of 6.5 years. Of the 20 subjects, 15 were males and five females, with a mean age of 27. Sixteen subjects reported to use computers every day, four subjects several times a week. Fourteen subjects reported familiarity with object-oriented systems development from courses, 11 subjects had such familiarity from systems development projects.

3.4 Design

The experiment employed a within-subjects factorial design, with the independent variables being interface type (linear vs. fisheye vs. overview+detail) and task type (essay vs. questions-answering). The experiment consisted of three sessions. In each session the 20 subjects used one interface to solve a task of each type. Each session lasted approximately one hour and 45 minutes, giving a total of 106 hours of experimental data. Tasks and interfaces were systematically varied and counterbalanced. We formed six groups based on permutations of the three interfaces. Using L to designate the linear, F the fisheye and O+D the overview+detail interface, these groups used the following orders of interfaces: L•F•O+D, L•O+D•F, F•L•O+D, F•O+D•L, O+D•L•F, O+D•F•L. Because six was not a divisor of the number of subjects, four groups comprised three subjects and two groups comprised four subjects. The tasks for these six groups were found by randomly choosing Latin squares such that the three interfaces and the three sessions had an approximately equal number of different tasks.

3.5 Reading Patterns and Usability Measures

One of the contributions of this paper is the description of reading patterns based on visualizations of reading activity described in section 4. Reading patterns are described in terms of reading modes and events. To ensure that modes and events were reliably detected, one of the authors first developed a classification of reading modes and events and applied it on all visualizations of reading activity. This happened blind to which interface the subjects had used. Afterwards, the other author classified a random sample of 20% of the visualizations of reading activity. The Pearson correlation between the authors' estimation of the duration of reading modes in essay tasks were between .96 and .99. For reading events, only the classification of one task differed. For question-answering tasks, the correlations were between .89 and .97. For the analysis in section 4, we used the classification of visualizations in the sample agreed upon by the authors. Visualizations not in the sample were adjusted to reflect the consensus among the authors.

To uncover the usability of the interfaces, we measured the following:

- *Grades* were given to all tasks. The answers were graded blind by the first author, i.e., without any knowledge of which subject had made the answer or with which interface the answer had been made. We used a five point grading scale, ranging from zero—a missing or completely wrong answer—to four—an outstanding and well-substantiated answer. For the question-answering tasks, grades were given according to how many aspects of the question the answer covered. A classification of the main ideas in the documents and important aspects of questions were developed to assist a systematic and uniform grading.
- *Incidental learning* was measured as the number of correct answers to incidental learning questions, resulting in a score from 0 to 6.
- *Task completion time* was used as the indicator of efficiency. All subjects' interactions with the interfaces were logged and the task completion times were derived from the data logged. For essay tasks, only the time spent reading is considered task completion time, leaving out the time spent writing the essay.
- *Satisfaction* was measured in three ways. After using each interface, subjects answered twelve questions about the perceived usability of the interface and their experiences with solving the tasks. After having used all three interfaces, subjects indicated which they preferred. Subjects also wrote comments about the interfaces after using each of them, and described why they preferred using one of the interfaces.

3.6 Procedure

The experiment took place in a lab without external disturbances. Two subjects participated at a time. Upon arriving, subjects filled out a questionnaire on background information and on their familiarity with object-oriented systems development. Then, subjects were trained until they felt confident in operating the interfaces. Training was supported by a two-page description of how to operate the interfaces. The subjects also completed three training tasks, which introduced them to the interfaces, and the question-answering and essay tasks. The mean time used to complete the training tasks was 35 minutes. After training, the subjects completed the first session of the experiment. Subjects returned the next day to the lab and completed the remaining two sessions.

The subjects received the tasks on sheets of paper, on which they also wrote the answers for the question-answering tasks. After finishing reading documents, the subjects proceeded immediately to the writing of essays, for which they received paper and pencil. The subjects were not allowed to take notes while reading the documents.

3.7 Analysis

The experimental design was expected to result in 20*3 solutions to the essay tasks, but one subject did not complete a task, and one solution was dropped from the analysis because of a time usage three interquartile ranges above the 75-quartile, leaving 58 solutions. The task completion time for that solution was 163 minutes, in comparison to the overall average of 42 minutes. For the question-answering tasks, the design should give 360 (20*3*6) answers, but one subject failed to complete a task, leaving 354 answers.

We analyzed the data by ANOVAs with interface type, task, session, and subject as factors. Essay tasks and question-answering tasks were analyzed separately.

4 Reading Patterns

4.1 Reading in Essay Tasks

4.1.1 Progression Maps and Reading Modes

We visualize each subject's reading activity for an essay task using progression maps. The progression maps show what parts of a document subjects have visible at which time in the reading process. Figure 2 shows an example of a progression map for an essay task. On the

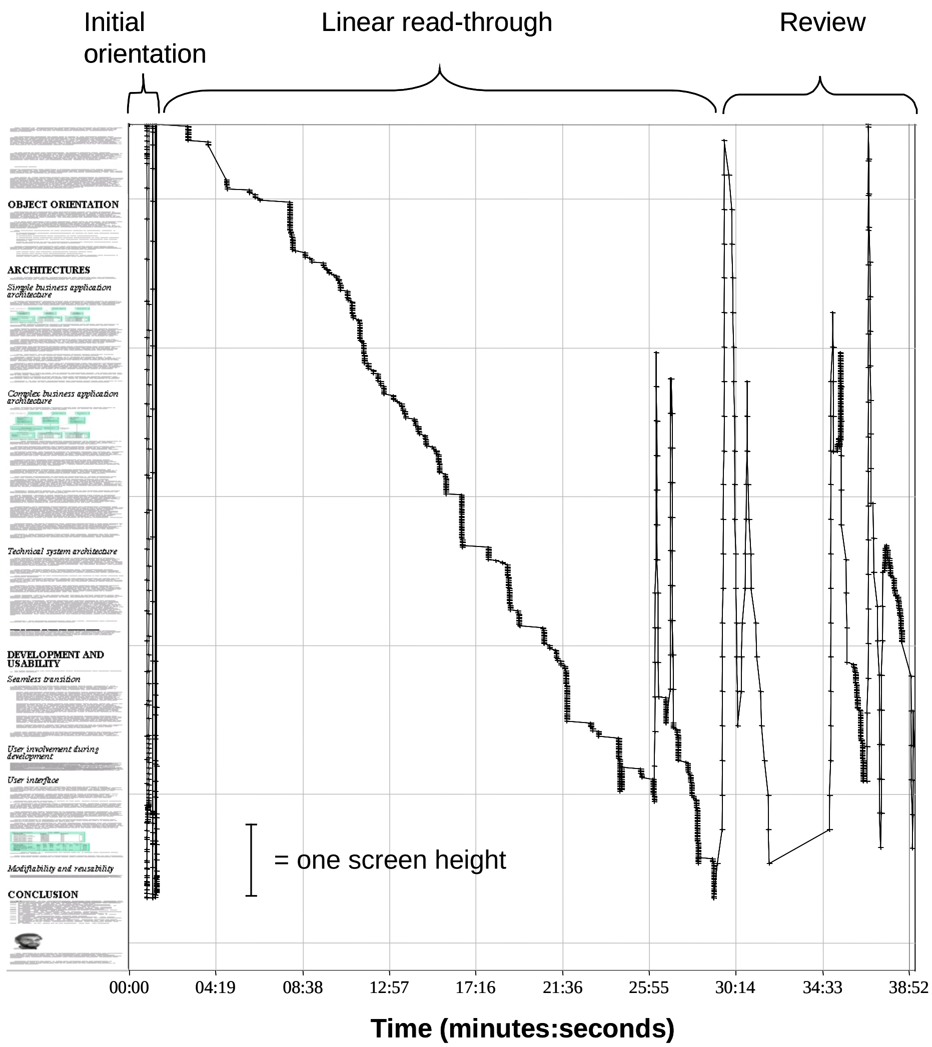


Figure 2—Progression map showing reading modes. This figure shows a progression map for a subject doing an essay task. The reading modes are indicated at the top of the figure. The horizontal axis shows time elapsed since the beginning of the task. The vertical axis shows the position in the document visible to the subject as the top-most position in the detail window. The vertical axis also shows an overview of the contents of the document. In the figure is indicated the height of one screen in the linear and overview+detail viewer. For the fisheye interface, subjects can see approximately twice as much.

progression maps, we identified three modes to describe how subjects read a document (see Figure 2). In the *initial orientation mode*, subjects navigated through the document in a non-linear fashion. We found this mode at the beginning of a task, if the subject attempted initial orientation. The initial orientation mode ends when subjects began reading linearly through the document from the beginning. In the *linear read-through mode,* subjects read through the document from the beginning to the end in a linear way, with occasional skips forwards and backwards. This mode ended when subjects began to navigate non-linearly through the document for more than one minute and do not return to continue the linear read-through. In the *review mode*, subjects looked again at what appears to be the most important sections in the document in a non-linear order. This mode was found at the end of a task. In every task we found an initial orientation mode. In 34 tasks we found an initial orientation mode and in 56 tasks a review mode.

Figure 3 shows the average duration of the three reading modes. We found significant differences in time spent in the modes for the initial orientation mode ($F[2,32]=3.38$, $p<.05$). In the fisheye interface, more time was spent in the initial orientation mode ($M=4.6$ min., $SD=5.5$) compared to the linear ($M=2.1$ min., $SD=3.2$, $F[1,32]=5.02$, $p<.05$) and the overview+detail interface ($M=2.0$ min., $SD=3.2$, $F[1,32]=5.11$, $p<.05$). A significant difference between the interfaces was also found in the time spent in the linear read-through mode, $F(2,32)=10.86$, $p<.001$. A linear contrast shows that subjects spend only two-thirds as long time with the fisheye interface in the read-through mode ($M=26.6$ min., $SD=16.2$) as with the other two interfaces (linear: $M=37.0$ min., $SD=10.6$, $F[1,32]=15.23$, $p<.001$; overview+detail: $M=37.5$ min., $SD=11.7$, $F[1,32]=17.25$, $p<.001$). For the review mode, we find no significant difference, $F[2,32]=1.48$, $p>.2$.

We made two further observations about the reading behaviour in the initial orientation mode. First, on the progression maps we repeatedly observed an orienting behaviour from the subjects that we call *flip-through*. In a flip-through, subjects scrolled through the entire document in less than 30 seconds (see Figure 2 for an example). Subjects did so at the beginning of an essay task. This behaviour seems similar to flipping through the pages in a book or a journal. We observed

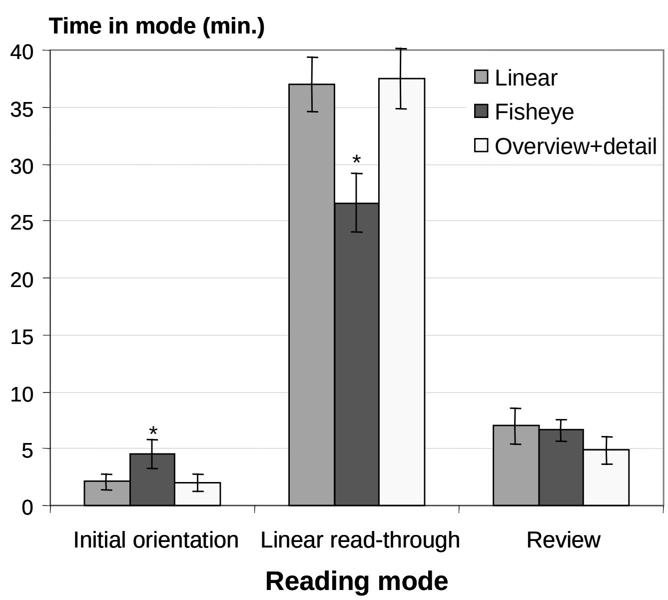


Figure 3—Time spent in reading modes. The figure shows the average time subjects spend in the three reading modes for each of the interfaces. Subjects without a certain mode were counted as spending zero minutes in that mode. An asterisk denotes a significant difference between interfaces. Error bars show the standard error of the mean.

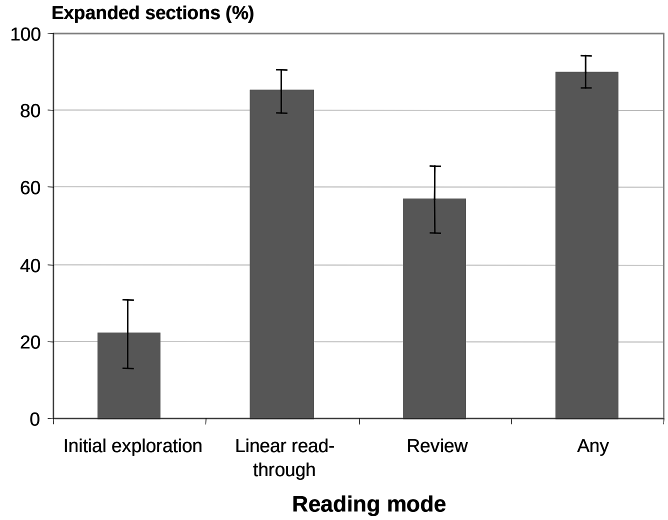


Figure 4—Percentage expanded sections in each reading mode. This figure shows the average number of sections that subjects expanded or that were kept expanded in the three reading modes. The rightmost bar shows the number of sections open in any reading mode. Error bars show the standard error of the mean.

flip-throughs in 30 out of the 59 essay tasks, with no difference between interfaces. Subjects may have used flip-throughs for obtaining an overview of the documents, a task that is notoriously difficult for electronic documents [O'Hara & Sellen 1997].

Second, we noticed that subjects during the initial orientation mode almost exclusively looked at the introduction and the conclusion of the paper, see Table 1.

4.1.2 Expansion and Collapsing of Sections in the Fisheye Interface

When using the fisheye interface, subjects on the average expanded 90% (SD=18) of the sections in a document, see Figure 4. Six subjects in one or more tasks expanded all sections at once by selecting the pop-up menu item 'expand all'; the rest of the subjects expanded sections by clicking with the mouse on the section. We also examined in what reading modes subjects expanded sections or kept previously expanded sections expanded. Our hypothesis was that the fisheye interface should support an overview-oriented reading style, meaning that subjects expanded sections primarily in the linear read-through mode. In the initial exploration mode, subjects expanded or kept expanded approximately one fourth (M=22%, SD=32) of the sections in the document. In the linear read-through mode, subjects expanded or kept expanded 85% (SD=24) of the sections. In the review mode, subjects expanded or kept expanded approximately

Document part	*Percentage tasks with parts visible*
Introduction and abstract	76%
Conclusion	41%
Other sections	18%
References and appendices	12%

Table 1—Document parts visible in the initial orientation mode. The table shows the percentage of the 34 tasks with an initial orientation mode where the document parts described in the left-most column are visible for more than one minute.

half (M=57%, SD=37) of the sections.

4.1.3 Visibility Maps

For all essay tasks, we also visualize reading activity by visibility maps. The maps were made by arbitrarily dividing the document into 100 parts of equal length. For each subject, we replayed the logged interaction and registered which parts were visible and for how long. Figure 5 shows an example of a visibility map for one essay task. To test the differences between interfaces revealed by casual inspection of the visibility maps, we compared the average percentage of the reading time spent in collapsed versus initially readable parts of the documents, see Figure 6. These maps, and the accompanying tests, reveal three interesting patterns about how long different parts of the documents were visible.

First, the visibility maps show that the relative duration for which different parts were visible differs between interfaces, see Figure 6. Here, we look at those parts of the documents that in the fisheye interface were initially readable. We found a significant difference between the duration these parts were visible between interfaces, $F[2,32]=35.2$, $p<.001$ (we used the arcsine transformation on the percentage values before running ANOVAs). In the fisheye-interface (M=13%, SD=4.4) the initially readable parts were visible for approximately 50% longer than in the linear interface (M=8%, SD=2.7, $F[1,32]=56.3$, $p<.001$) and the overview+detail interface (M=9%, SD=3.2, $F[1,32]=48.8$, $p<.001$). Similarly, we find a difference between interfaces in how long parts, which in the fisheye interface were initially collapsed, were visible, $F[2,32]=36.0$, $p<.001$. Linear contrasts show that in the fisheye interface (M=5%, SD=1.5) these parts were visible shorter compared to the other two interfaces (linear: M=7%, SD=1.9, $F[1,32]=60.2$, $p<.001$, overview+detail: M=7%, SD=1.3, $F[1,32]=46.9$, $p<.001$).

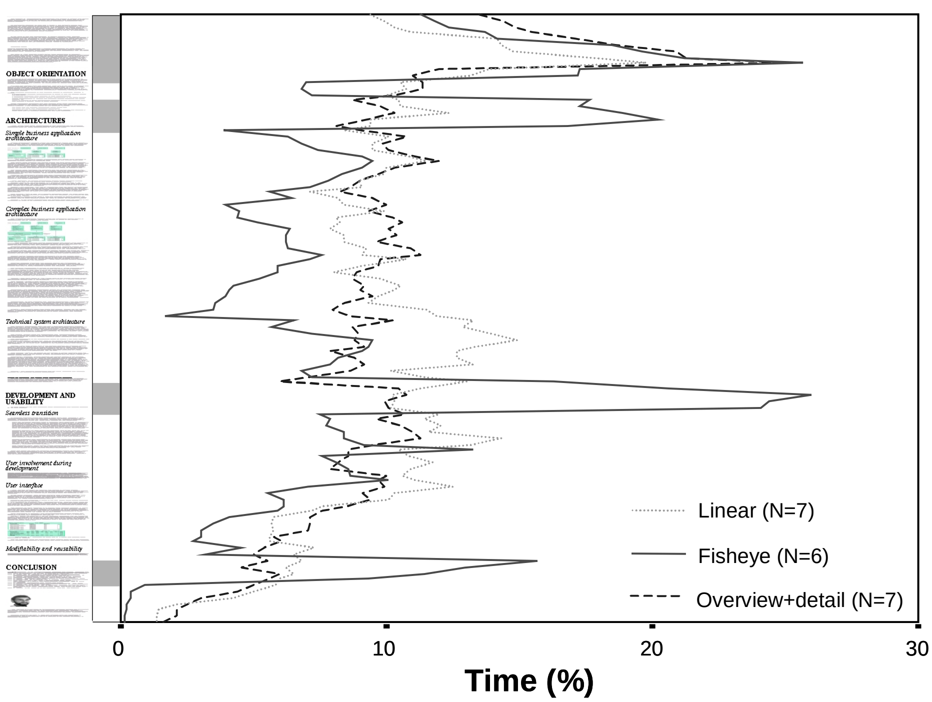


Figure 5—Visibility map for one of the three essay tasks. The horizontal axis shows the average time a part of the document is visible for each interface. The vertical axis shows position in the document, as indicated by the overview of the document. The grey squares along the vertical axis indicate parts of the document that were initially readable in the fisheye interface.

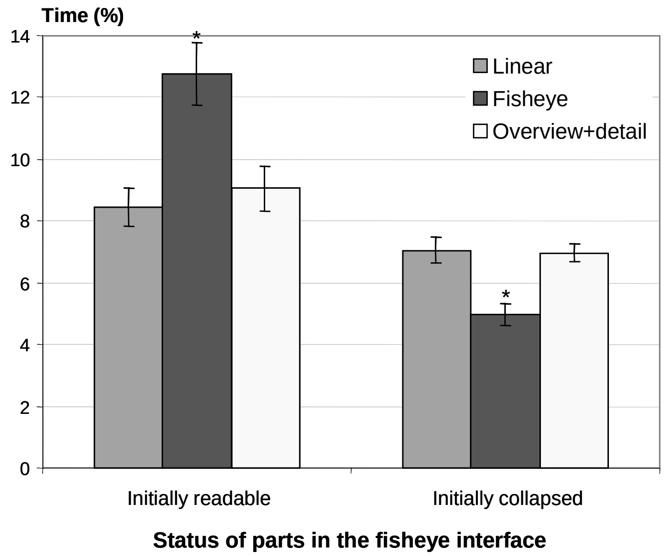


Figure 6—Time spent in parts of the documents that are initially readable vs. initially collapsed in the fisheye interface. This figure shows the average time spent in those parts of the documents that in the fisheye interface are either initially readable or initially collapsed. An asterisk indicates a significant difference between the interfaces. Error bars show the standard error of the mean.

Second, in the overview+detail interface and the linear interface, subjects have sections visible a comparable length of time.

Third, for the linear and overview+detail interface we find a difference between how long certain parts of the document are visible. The time spent in parts of the document that in the fisheye interface are initially readable, is longer with the overview+detail and linear interface compared to the time spent in parts that are initially collapsed in the fisheye interface. This suggests that the algorithm for the fisheye interface chooses sections to be initially readable that subjects spend relatively long time reading.

4.2 Reading in Question-answering Tasks

For question-answering tasks, we visualize reading activity for each subjects' answer to each of the six questions on a progression map. To analyse these maps, we use a notion of *targets* in the documents, of reading events called *first contact*, and reading modes called *target reading* and *further explorations* (see Figure 7). A *target* is a part of the document in which an answer to the

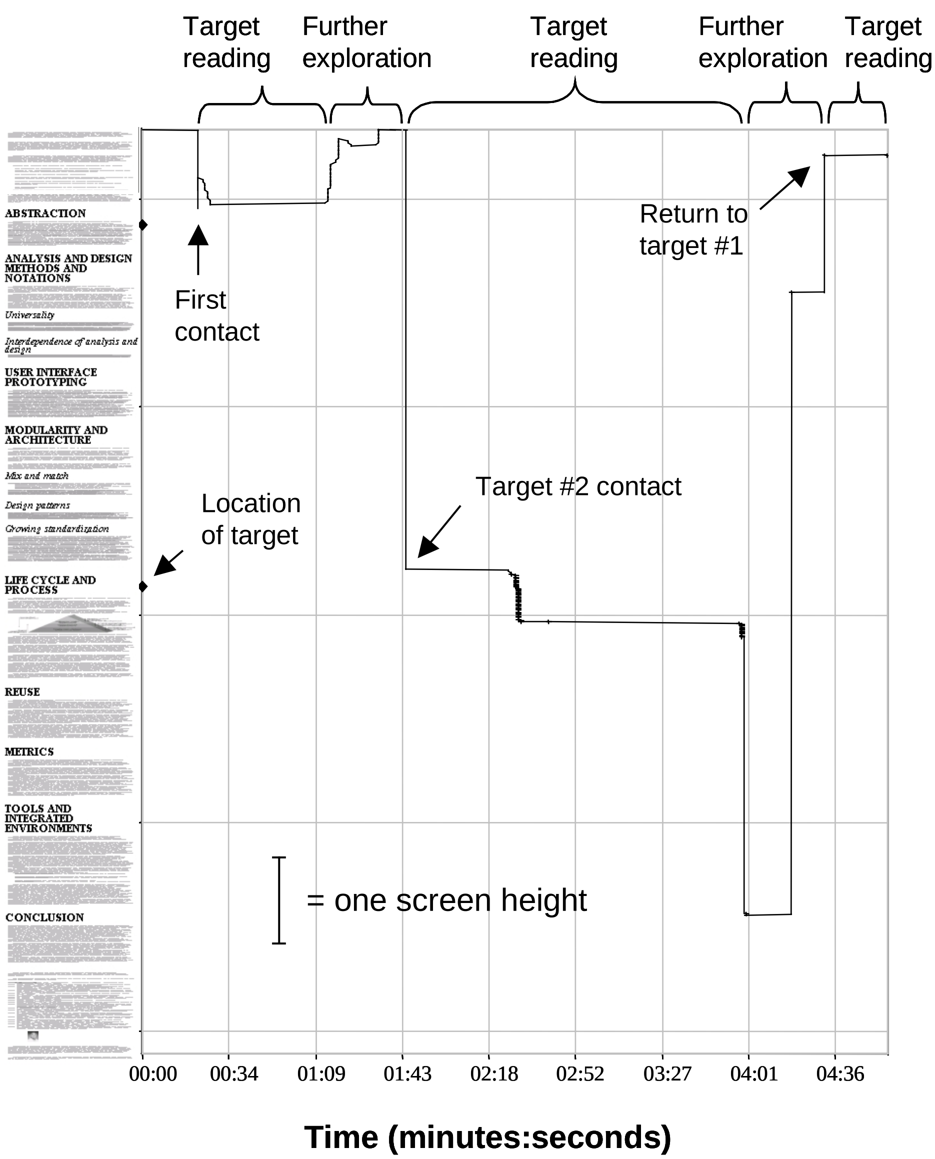


Figure 7—An example of a progression map for one question in a question-answering task. The reading modes are indicated at the top of the figure. The horizontal axis shows time elapsed since the beginning of the task. The vertical axis shows the position in the document visible to the subject as the top-most position in the detail window. The vertical axis also shows an overview of the contents of the document. In addition, targets and reading events are shown on the map. In the figure is indicated the height of one screen in the linear and overview+detail viewer. For the fisheye interface, subjects can see approximately twice as much.

current question can be found. In two questions for each of the question-answering tasks, the answer to the question (or a substantial part of it) can be found in more than one place. On the progression maps the target is shown as a point, but obviously both text right before and after the target point are important. Thus we consider a subject to do *target reading* as long as the target is visible in the browser window and when the target is less than half a screen-length above the top of the detail window. For the fisheye interface the length of the document parts visible in the window varies, because contents in the visible area may be collapsed. For this interface, we therefore used one screen length above the top of the window to delimit the target area.

4.2.1 First Contact with a Target and Direct Jumping

First contact is the moment when any target for the first time becomes visible in the detail part of the interface. To be considered a first contact, the target area must be visible for at least 20 seconds. If the target is at the beginning of the document, the subject might have been reading the task description. Therefore we begin to look for a first contact after 10 seconds. Figure 8 shows the average time passed from the beginning of the task to the moment where subjects make first contact. We find no difference between interfaces in how fast subjects made first contact, $F[2,313]=.341$, $p>.5$.

Figure 8 also shows that the number of targets found differed between interfaces, $F[2,313]=6.97$, $p<.001$. A linear contrast between interfaces suggest that 10% more targets were located in the overview+detail interface ($M=1.16$, $SD=.44$) compared to the linear interface ($M=1.05$, $SD=.21$), $F[1,313]=13.2$, $p<.001$. Note that only one third of the tasks contains multiple targets. The difference in number of targets located suggests that subjects keep exploring the document in the overview+detail condition, even when a satisfactory answer has been found.

In some tasks, subjects went directly to a target by clicking on the overview pane, a *direct jump*. In 54 out of the 120 tasks solved with the overview+detail interface, subjects made first contact this way. Subjects therefore seem able to relate the information on the overview to the questions. In 13 tasks, subjects return to an already visited target by a direct jump. Since subjects only return in 44 of the tasks solved with the overview+detail interface, this account for 30% of the returns. The use of the overview pane for returning to targets suggests that subjects remember the position of previously visited parts of the document on the overview pane.

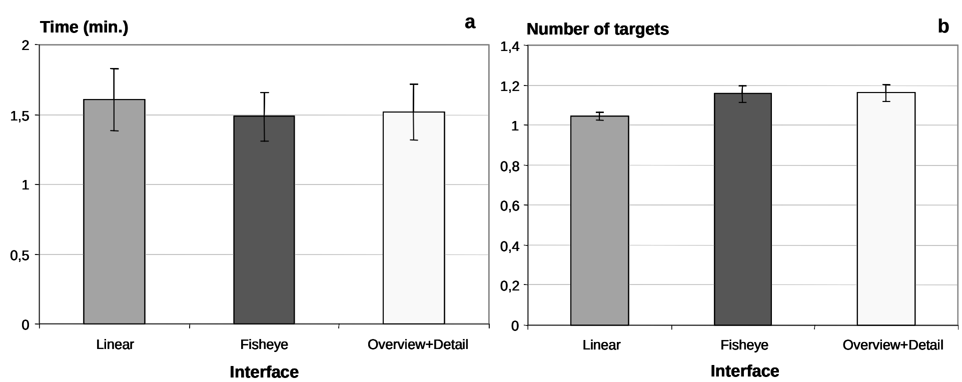


Figure 8—Time to first contact and number of targets reached. Panel **a** shows the average time to first contact in the three interfaces. Panel **b** shows the average number of targets found in the three interfaces. Error bars show the standard error of the mean. This figure only includes tasks in which one or more targets are reached (N=335).

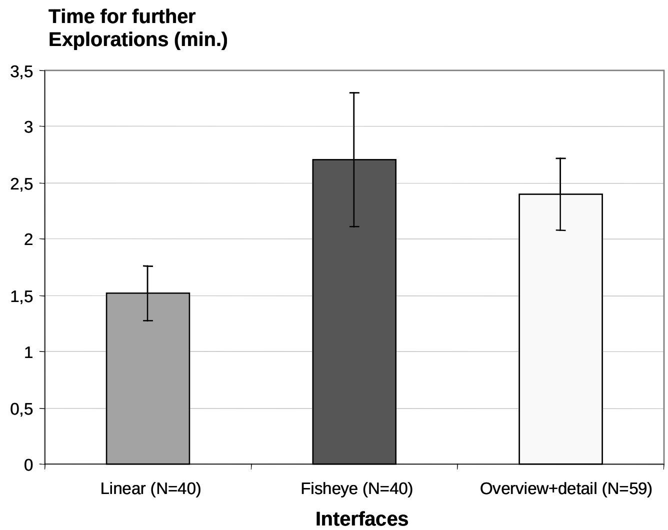


Figure 9—Further explorations. This figure shows the frequency of tasks with one or more further explorations and the average duration of further explorations in the three interfaces. Error bars show the standard error of the mean.

4.2.2 Further Explorations

When subjects, after having made first contact, stop target reading and navigate to a non-target area, we say they make *further explorations*. We do not consider it a further exploration if the subject navigated directly to another target. Further explorations had to last more than 10 seconds.

Figure 9 shows the average number of question-answering tasks in which subjects explore the document further. The number of further explorations were significantly different between interfaces, χ^2 [2, N=354]=7.59, p<.05. Subjects explored the document further in 48% more tasks in the overview+detail interface compared to the linear and the fisheye interface. Figure 9 also shows that subjects explore the document for different lengths of time, F[2,313]= 3.87, p<.05. Compared to the linear interface (M=1.6 min., SD=.26), significantly more time is used exploring the document in overview+detail interface (M=2.4, SD=.32, F[1,313]=7.46, p<.01).

Table 2 shows the different actions with which subjects started further explorations. Note that this table shows the total number of further explorations, not just the number of tasks in which a further exploration occurs (as in Figure 9). In the overview+detail interface, subjects clicked on the overview pane to navigate to the area clicked on. The progression maps show that subjects used this feature to begin further exploration twenty-two (26%) times.

Interface	*Highlight*	*Scroll up*	*Scroll down*	*Jump on overview*
Linear (N=50)	11 (22 %)	16 (32 %)	23 (46 %)	NA
Fisheye (N=53)	10 (19 %)	18 (34 %)	25 (47 %)	NA
Overview+detail (N=84)	8 (10 %)	22 (26 %)	32 (38 %)	22 (26 %)

Table 2—How do Further Explorations Begin? The table shows the number (and percentage in parenthesis) of further explorations started by the actions shown in the top row.

5 Usability Measures

This section presents the differences in measures of usability between interfaces. We use the reading patterns presented in the previous section to explain these differences.

5.1 Grades and Incidental Learning

Figure 10 shows the average grade and incidental-learning score for the three interfaces. For essay tasks, we found a significant influence of interface on the average grade obtained, $F[2,32]=4.16$, $p<.05$. Linear contrasts show that tasks solved with the overview+detail interface ($M=2.47$, $SD=.84$) on average got half a grade higher compared to the linear ($M=2.00$, $SD=.86$, $F[1,32]=5.26$, $p<.05$) and the fisheye interface ($M=1.95$, $SD=.78$, $F[1,32]=7.10$, $p<.05$). Based on the reading patterns, we have no direct explanation for this finding. However, the question-answering tasks suggest that subjects are able to use the overview pane to navigate and that they remember the position of information on the pane. In addition, we speculate that the overview pane may indirectly have helped subjects to organize and recall text.

The number of correctly answered incidental learning questions differed significantly between interfaces, $F[2,32]=6.80$, $p<.01$. Subjects correctly answered fewer questions in the fisheye interface ($M=3.42$, $SD=1.22$) compared to the linear interface ($M=4.20$, $SD=1.24$, $F[1,32]=8.22$, $p<.01$) and the overview+detail interface ($M=4.58$, $SD=1.22$, $F[1,32]=11.83$, $p<0.01$). On the average, around one question less was correctly answered when subjects used the fisheye interface. The visibility maps suggest that subjects pay less attention to initially collapsed sections, thereby missing information for some incidental learning questions. The overview-oriented reading style of the fisheye interface apparent from the analysis of reading modes, suggests a similar reason for subjects' low incidental-learning score.

For question-answering tasks, we found no difference between interfaces for the grades given to the tasks, $F[2,313]=.18$, $p>.5$.

5.2 Task Completion Time

Figure 11 shows the task completion time for essay and question-answering tasks. For essay tasks we find a difference in task completion time, $F[2,32]=4.92$, $p<.05$. A linear contrast analysis shows that the fisheye interface ($M=37.4$ min., $SD=12.4$) were approximately 16% faster than the

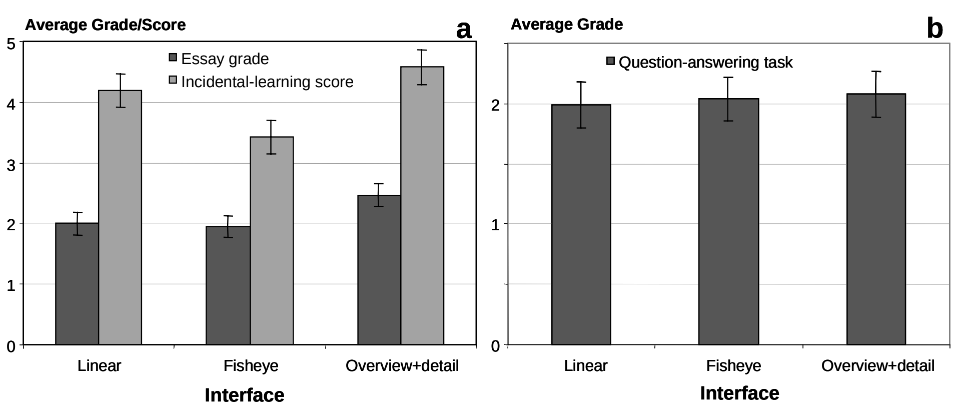


Figure 10—Grades and incidental learning. Panel **a** shows for the essays tasks the average incidental learning score and grade for the three interfaces (N=58). Panel **b** shows for the questions-answering tasks the average grade for each question (N=354). Error bars show the standard error of the mean.

linear interface (M=44.4 min., SD=11.9, F[1,32]=8.13, p<.01) and the overview+detail interface (M=44.5 min., SD=12.2, F[1,32]=6.51, p<.05). The reading patterns explain why subjects are faster with the fisheye interface. The overview-oriented reading style and the short time subjects look at initially collapsed sections appear as the main reasons. Note that subjects using the fisheye interface have to expand most of the sections in the document: what is different from the other interfaces is the duration these sections are visible.

For the question-answering tasks we also found a significant difference between interfaces, F[2,313]=4.235, p<.05. The overview+detail interface (M=7.1 min., SD=4.1) were 20% slower compared to the linear interface (M=5.9 min., SD=3.5, F[1,313]=8.33, p<.01). Note, as explained in section 3.2 we imposed no time limit on the subjects work with the tasks. According to the reading patterns, this time difference is not due to difficulty in locating a target. However, in the reading patterns further explorations are more frequent and last longer in the overview+detail interface. Another indicator of this is that the number of targets found with the overview+detail interface is higher than in the other interfaces. Interestingly, subjects initiated many further explorations by clicking on the overview pane.

5.3 Satisfaction

Nineteen of the subjects preferred using the overview+detail interface; one subject preferred the linear interface. In their motivation for preferring the overview+detail interface, 10 subjects mentioned the overview of the documents structure and headings as an important reason; six subjects mentioned that the overview+detail interface support easy navigation; and five subjects liked that highlighted words show up in the overview pane. Fourteen subjects mentioned that they found it hard to overview the document using the linear interface. With respect to the fisheye interface, nine subjects commented that they did not like to depend on an algorithm to determine which parts of the document should be readable. Subjects were divided as to whether the fisheye made it easier (N=5) or harder (N=2) to get an overview of an article. Figure 12 shows the subjects' answers to the questionnaires received after using each of the interfaces. We compared interfaces using paired t-tests with a Bonferroni-adjustment of 0.05/12*3≈.0013. The overview+detail interface scored higher than the two other interfaces on satisfaction questions about overall satisfaction, and on the dimensions terrible-wonderful and frustrating-pleasant. Subjects scored the fisheye interface lower compared to the overview+detail interface on the dimension confusing-clear. Subjects also scored the overview+detail interface higher compared to

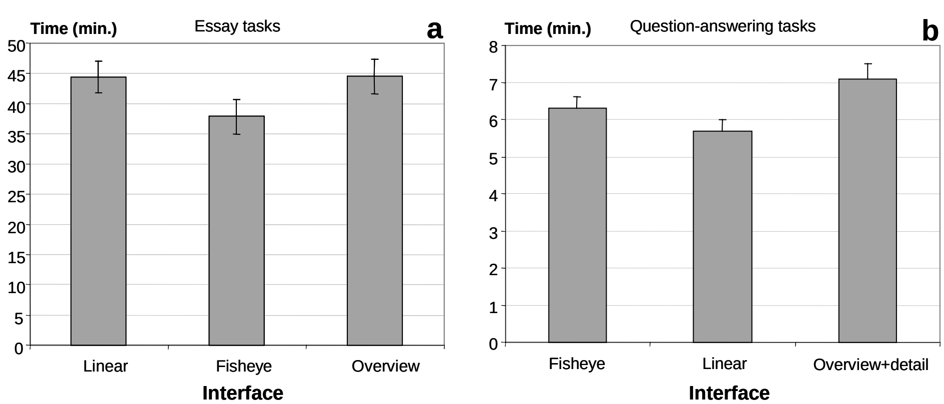


Figure 11—Task completion time. Panel **a** shows the average task completion time of essay task grade for the three interfaces. Panel **b** shows the average task completion time for question-answering tasks. Error bars show the standard error of the mean.

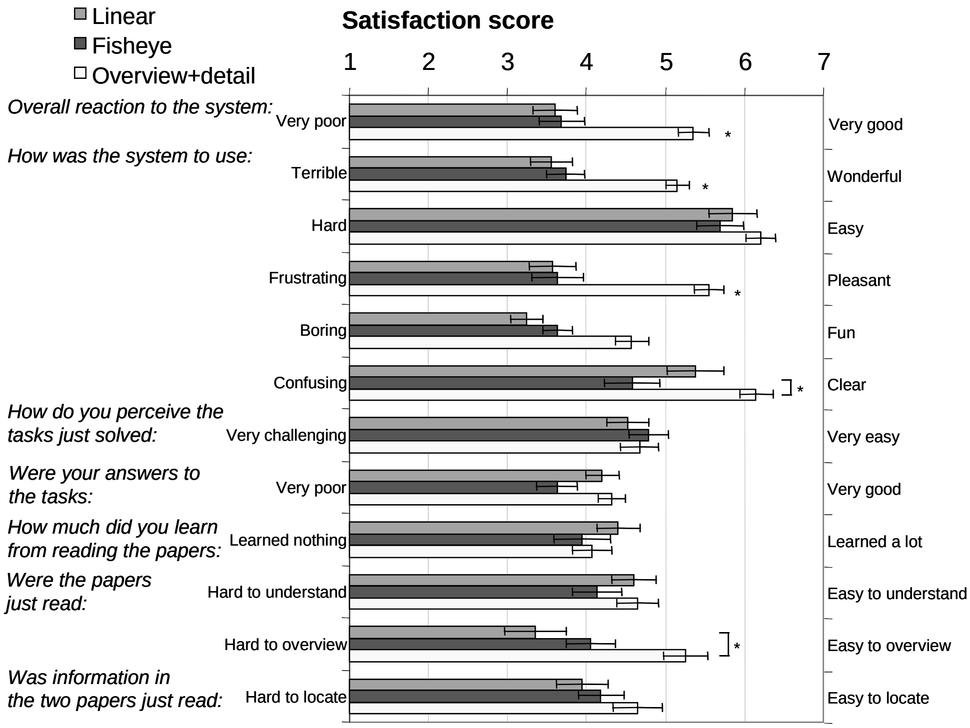


Figure 12—Satisfaction. To the left, in italics, is shown the questions subjects were given. Negative and positive concepts on a seven-point semantic differential are shown to the left and right of the chart. Low scores were given to the negative concept of the differential scale. The bars show the average satisfaction scores for the three interfaces. An asterisk denotes a significant difference using a Bonferroni adjustment of .0013.

the linear interface on the question whether the documents were easy or hard to overview. We found no difference for the questions intended to investigate whether the subjects' perception of their tasks differed between interfaces (question 7 and 8 in Figure 4).

The satisfaction with the overview is supported by several of the reading patterns. The overview pane support jumping directly to targets; it helps returning to previously visited parts of the document; and it invites to and supports further explorations. Subjects using the fisheye interface depend extensively on the algorithm that determines which sections to collapse initially, even though subjects do not trust this algorithm.

6 Discussion

The overview+detail interface was slow for question-answering tasks. Further explorations were more frequent in the overview+detail interface compared to the two other interfaces and were often initiated by clicking on the overview pane. These observations support the explanation previously proposed [Hornbæk & Frøkjær 2001] that the overview pane grabs subjects' attention, and thereby lead them to explorations that strictly speaking are unnecessary. Seemingly, this happens because of the visual appearance of the overview and because of the navigation possibilities afforded by the ability to click the overview pane. The reading process we observed thus seems more unpredictable and shaped by situation-dependent inspiration compared to the description offered by Guthrie's model [Guthrie 1998] of locating information in documents, which suggests a rational, goal-oriented process. Another point is that the overview in our experiment slow down task completion time. This result is in contrast to the expectations raised in previous work (see section 2) and our hypothesis (see section 3) about the usability of overview+detail interfaces. Our results are similar to the empirical results of [Dee-Lucas & Larkin 1995; Hornbæk et al. 2001] who found that overviews may lead to higher task completion times.

However, the overview+detail interface leads to higher quality essays and subjects strongly preferred this interface. We found several indicators why this happened: the overview support navigation, invite explorations, and support jumping directly to previously read text. Thus, we think designers should be well advised to use overview+detail interfaces for electronic documents.

For essay tasks, the fisheye interface was approximately 16% faster. Subjects opened almost all collapsed sections, but spent less time on the initially collapsed sections compared to the other interfaces. Subjects also used more time in the initial orientation mode and less time in the linear read-through mode. On one hand, our hypothesis about the fisheye interface was confirmed: the fisheye interface shortens navigation time and supports an overview-oriented reading style. On the other hand, we were surprised that the initial status of sections influenced the duration they were visible as strongly as observed. We suspect that subjects assume that the contents of the initially collapsed sections are not important, independently of what they read in the sections. This behaviour is akin to premature cognitive commitment [Langer 1991], where humans commit themselves to one view on or use of information and at a later time fail to reconsider their commitment. This premature commitment comes about even though many subjects expressed a lack of trust in the algorithm. Thus, fisheye interfaces may fundamentally change the way subjects perceive and interact with documents. The lower task completion time might account for the lower incidental learning scores obtained by subjects using the fisheye interface. Together, these observations suggest that fisheye interfaces should be used mainly for time-critical tasks and for tasks where a detailed understanding of the document is not the main aim, for example in relevance judgements such as judging whether it is worthwhile to download or thoroughly read a document.

The linear interface is in many ways clearly inferior with respect to usability compared to the two other interfaces. We recommend that designers rely less on this interface type and use the overview+detail and, in special cases, the fisheye interface.

Our visualizations of reading patterns suggest four interesting observations in addition to those mentioned above. First, the flip-through behaviour suggests that subjects develop techniques for coping with the low tangibility of electronic documents [Hansen & Haas 1988; O'Hara & Sellen 1997]: flipping through the document might give subjects an initial indication of the length, structure and key elements of the document. Many subjects seem to like doing a flip-through to set the scene for a more careful reading of the documents. Second, in question-answering tasks the reading patterns show how subjects used the overview-pane to navigate back to previously visited targets. This indicates that the overview pane supports memorizing important document positions, perhaps in a way analog to the way readers remember the position of information in paper documents [Rothkopf 1971]. Third, our observations on reading patterns confirm and extend previous research on reading, e.g. Bishop [1999]. Non-linear navigation occurred extensively, but mostly at the beginning and end of the reading activity. We also found, similarly to [Foltz 1996], that most of the reading time consisted of linear reading through the document. Fourth, large differences between individuals in reading strategies were also found, as do for example Goldman & Saul [1990]. However, in this paper we concentrated on examining the influence of interface on reading patterns.

Concerning techniques for studying reading patterns, progression maps offer an intermediate analysis tool of user behaviour in reading electronic documents, between coarse measures, such as task completion time, and fine detail analysis, such as eye-tracking analysis. The most important limitation of our technique is that we only register the visible parts of the documents, not what subjects actually looked at.

In relation to the aims set forth in the introduction to this paper, we have investigated both the usability and the reading process in visualizations of electronic documents. To follow up this investigation, we suggest three areas of further research. First, we need to improve visualizations of electronic documents. The algorithm for making the fisheye interface may be improved based on our descriptions of reading patterns. The overview pane may also benefit from more information-rich semantic zooming. Second, we need a more thorough study of reading activity during actual work as performed by subjects who have gained full familiarity with the experimental interfaces. Third, we need some better theories of how attention is shaped by visualizations. The role of the overview in triggering further explorations is not well described by theories of information visualization we are aware of.

7 Conclusion

In an experiment, we compared three interfaces for electronic documents. Two of the interfaces were based on overview+detail and fisheye visualizations; the third was a linear interface that served as a baseline. Subjects in the experiment answered questions and wrote essays about scientific documents. In an attempt to better understand how the interfaces supported reading and to understand the differences in measures of usability between interfaces, we created visualizations of subjects' reading activity by two kinds of maps. Progression maps were used to depict how the reading progressed; visibility maps were used to compare the average time different parts of the document were visible. From these visualizations we describe how the interfaces shape subjects' reading patterns.

Subjects clearly preferred the overview+detail interface, especially because of the overview gained and the ease of navigation. With this interface, essays received a higher grade. For question-answering tasks, the progression maps show that subjects with the overview+detail interface explore the document more often than with the other interfaces. Consequently, subjects use longer time answering questions. The visibility maps reveal that subjects with the fisheye interface have visible for less time the parts of the document that are not initially readable. With the fisheye interface, subjects also read the documents using an overview-oriented reading style. Therefore subjects read faster with this interface, but display lower incidental learning.

As for the practical problem of using visualizations to support reading, visualization interfaces improve the usability of electronic documents. However, visualizations also change how subjects read documents. The most common interface in practical use, the linear interface, was inferior on most usability aspects compared to the other two interfaces. Visualizations are thus recommended to developers as usable interfaces for electronic documents. For researchers, further improvements of visualizations of electronic documents are feasible, as are use of progression and visibility maps to study and improve reading activity.

8 Acknowledgements

We thank Peter Naur and Morten Hertzum for comprehensive comments on a draft of this paper.

9 References

Anderson, T. H. & Armbruster, B. B. (1982). Reader and Text-Studying Strategies. In *Reading Expository Material.* Otto, W. and White, S. New York, NY: Academic Press. 219-242.

Bazerman, C. 1988. *Shaping Written Knowledge.* Madison, WI: The University of Wisconsin Press

Bishop, A. P. 1995. Working Towards an Understanding of Digital Library Use. *D-Lib Magazine*, October 1995, http://www.dlib.org/dlib/october95/10bishop.html.

Bishop, A. P. 1999. Document structure and digital libraries: how researchers mobilize information in journal articles. *Information Processing and Management, 35,* 255-279.

Boguraev, B., Kennedy, C., Bellamy, R., Brawer, S., & Wong, Y. Y. 1998. Dynamic presentation of document content for rapid on-line skimming. In *AAAI Spring 1998 Symposium on Intelligent Text Summarization,* AAAI Press, Stanford CA, 109-118.

Bradow, R., Mitze, K., and Rau, L. F. 1995. Automatic condensation of electronic publications by sentence selection. *Information Processing & Management, 31,* 5, 675-685.

Byrd, D. 1999. A scrollbar-based visualization for document navigation. In *Proceedings of Digital libraries '99*, (DL'99, Berkeley CA 11-14 Aug.), ACM Press, New York, NY, 122-129.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. 1999. The tangled web we wove: a taskonomy of WWW use. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI '99, Pittsburgh PA, May 15-20), ACM Press, New York, NY, 544-551.

Chen, C. and Rada, R. 1996. Interacting with Hypertext: A Meta-Analysis of Experimental Studies. *Human-Computer Interaction, 11,* 2, 125-156.

Dee-Lucas, D. and Larkin, J. H. 1995. Learning from electronic texts: Effects of Interactive Overview for Information Access. *Cognition and Instruction, 13,* 3, 431-468.

Dillon, A. 1994. *Designing Usable Electronic Text.* London: Taylor & Francis.

Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., and Lochbaum, C. C. 1989. Formative Design-Evaluation of SuperBook. *ACM Transaction on Information Systems, 7*, 1, 30-57.

Eick, S. G., Steffen, J. L., and Sumner, E. E. 1992. Seesoft-A Tool for Visualizing Line Oriented Software Statistics. *IEEE Transactions on Software Engineering, 18,* 11, 957-968.

Foltz, P. W. (1996). Comprehension, Coherence and Strategies in Hypertext and Linear Text. In *Hypertext and Cognition.* Rouet, J. F., Levonen, J. J., Dillon, Andrew P., and Spiro, R. J. New Jersey: Lawrence Erlbaum Associates. 109-136.

Furnas, G. W. 1986. Generalized fisheye views. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI '86, Boston MA, Apr.), ACM Press, New York, NY, 16-23.

Ginsburg, A., Marks, J., & Shieber, S. 1996. A Viewer for Postscript Documents. In *ACM Symposium on User Interface Technology and Software* (UIST'96, Seattle, WA, 6-8 Nov.), ACM Press, New York, NY, 31-32.

Goldman, S. R. and Saul, E. U. 1990. Flexibility in text processing. *Learning and Individual Differences, 2*, 2, 181-219.

Graham, J. 1999. The reader's helper: a personalized document reading environment. In *Proceeding of ACM SIGCHI Conference on Human Factors in Computing Systems* (CHI'99, Pittsburgh PA, 15-20 May.), ACM Press, New York, NY, 481-488.

Guthrie, J. 1998. Locating information in documents: examination of a cognitive model. *Reading Research Quarterly, 23,* 178-199.

Hansen, W. J. and Haas, C. 1988. Reading and writing with computers: a framework for explaining differences in performance. *Communications of the ACM, 31,* 9, 1080-1089.

Hertzum, M., Lalmas, M., & Frøkjær, E. 2001. How are searching and reading intertwined during retrieval from hierarchically structured documents? In *Proceedings of the IFIP TC 13International Conference on Human-Computer Interaction (*INTERACT '01, Tokyo, Japan, July 9-13), IOS Press, Amsterdam, 537-544.

Holmquist, L. E. 1997. Focus+Context Visualization with Flip Zooming and the Zoom Browser. In *CHI '97 Extended Abstracts,* (CHI'97, Atlanta GA, Mar.), ACM Press, New York, NY, 263-264.

Hornbæk, K., Bederson, B., and Plaisant, C. 2001. Overview+Detail and Zoomable User Interfaces: An Evaluation of Navigation Patterns and Usability. *HCIL technical report #2001-11,* ftp://ftp.cs.umd.edu/pub/hcil/Reports-Abstracts-Bibliography/2001-11html/2001-11.htm. (under review for ACM Transactions on Computer-Human Interaction).

Hornbæk, K. & Frøkjær, E. 2001. Reading Electronic Documents: The Usability of Linear, Fisheye, and Overview+Detail Interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2001, Seattle, WA, Mar. 31- Apr. 5). ACM Press, New York, NY, 293-300.

Horney, M. and Anderson-Inman, L. 1994. The ElectroText Project: Hypertext Reading Patterns of Middle School Students. *Journal of Educational Multimedia and Hypermedia, 3,* 1, 71-91.

Kaugars, K. 1998. Integrated multi scale text retrieval visualization. In *CHI '98 Summary* (CHI'97, Los Angeles CA, Apr.), ACM Press, New York, NY, 307-308.

Kupiec, J., Pedersen, J., & Chen, F. 1995. A trainable document summarizer. In *Proceedings of Annual ACM Conference on Research and Development in Information Retrieval* (SIGIR'95, Seattle WA, July), ACM Press, New York, NY, 68-73.

Langer, E. J. 1991. *Mindfulness.* London: Harvill.

Mills, C. B. and Weldon, L. J. 1987. Reading text from computer screens. *ACM Computing Surveys*, *19,* 4, 329-357.

Muter, P. (1996). Interface Design and Optimization of Reading of Continous Text. In *Cognitive Aspects of Electronic Text Processing.* van Oostendorp, H. and de Mul, S. Norwood, NJ: Ablex.

O'Hara, K. & Sellen, A. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI'97, Atlanta, GA, Mar.), ACM Press, New York, NY, 335-342.

Páez, L. B., da Silva-Fh., J. B., & Marchionini, G. 1996. Disorientation in Electronic Environments: A Study of Hypertext and Continuous Zooming Interfaces. In *Proceedings of the 59th Annual Meeting of the American Society for Information Science* (ASIS '96, Baltimore, MD, Oct. 19-24). Harding, S. Ed., 58-66.

Plaisant, C., Carr, D., and Shneiderman, B. 1995. Image browsers: Taxonomy, guidelines, and informal specifications. *IEEE Software, 12*, 2, 21-32.

Robertson, G. G. & Mackinlay, J. D. 1993. The document lens. In *Proceedings of ACM Symposium on User Interface Technology and Software* (UIST'93, Atlanta GA, Nov.), ACM Press, New York, NY, 101-108.

Rothkopf, E. 1971. Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behaviour, 10,* 608-613.

Schriver, K. 1997. *Dynamics in Document Presentation.* New York/London: John Wiley & Sons.

Sellen, A. & Harper, R. 1997. Paper as an analytic resource for the design of new technologies. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI'97, Atlanta GA, Mar.), ACM Press, New York, NY, 319-326.

Wright, P. (1987). Reading and Writing for Electronic Journals. In *Executive Control Processes in Reading.* Britton, B. K. and Glynn, S. M. Hillsdale, NJ: Lawrence Erlbaum Associates. 23-55.

Paper 6—Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing

E. Frøkjær & K. Hornbæk (2001), 'Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing', 8 pages, to appear in *Proceedings of HF2002/OzCHI*, Melbourne, Australia, 25th-28th November 2002.

Metaphors of Human Thinking in HCI: Habit, Stream of Thought, Awareness, Utterance, and Knowing

Erik Frøkjær & Kasper Hornbæk*

Department of Computing,
University of Copenhagen,
Copenhagen, Denmark
+45 35321456
{erikf, kash}@diku.dk

ABSTRACT

Understanding human thinking is crucial in the design and evaluation of human-computer interaction. Inspired by introspective psychology, we present five metaphors of human thinking. The aim of the metaphors is to help designers to consider important traits of human thinking when designing. The metaphors capture aspects of human thinking virtually absent in recent years of the CHI Conference Proceedings. As an example of the utility of the metaphors, we show how a selection of good and poor user interfaces can be appreciated in terms of the metaphors. The metaphors are also used to reinterpret central notions in human-computer interaction, such as consistency and information scent, in terms of human thinking. Further, we suggest the metaphors be used for evaluating interfaces.

Keywords

Human thinking, habit, automaticity, stream of thought, consciousness, awareness, attention, association, utterance, knowing, introspective psychology, metaphors, design guidelines, evaluation

INTRODUCTION

We present several metaphors related to the human thinking activity, and show by examples how the metaphors may serve to clarify aspects of designs of human-computer interaction (HCI).

For some years our research and teaching in human-computer interaction have been inspired by William James's and Peter Naur's descriptions of human thinking [20,23-26]. Similar descriptions along with many brilliant design discussions have lately been introduced to HCI in Jef Raskin's book *The Humane Interface* [32]. Naur's and Raskin's work are complementary to most psychology used in HCI, but is supported by extensive evidence from classic introspective psychology [20], and from experimental psychology and neurology [1,2]. Several of the aspects of human thinking described in this work are of critical importance to human-computer interaction: (1) the role of habit in most of our thought activity and behaviour—physical habits, automaticity, all linguistic activity, habits of reasoning; (2) the human experience of a stream of thought—the continuity of our thinking, the richness and wholeness of a person's mental objects, the dynamics of thought; (3) our awareness—shaped through a focus of attention, the fringes of mental objects, association, and reasoning; (4) the incompleteness of utterances in relation to the thinking underlying them and the ephemeral nature of those utterances; and (5) knowing—human knowing is always under construction and incomplete.

In this paper we present five metaphors of human thinking that cover the phenomena mentioned above. The contribution of the metaphors is threefold. First, the metaphors introduce a clear and recognizable way of talking about human thinking which we find absent in recent CHI Conference Proceedings. Second, we use the metaphors to analyse commonly available user interfaces. This shows the utility of the metaphors in recognizing and exploiting important characteristics of human thinking. In addition we show how central notions in HCI can be understood in terms of the metaphors, which we claim lead to a gain in clarity and immediate understandability of these notions. Third, we suggest further application of the metaphors to user interface design and evaluation.

In the next section, we present the metaphors and show how they describe aspects of human thinking crucial to HCI. Then we show how the metaphors can be used to describe important phenomena in HCI. Finally, we discuss some limitations in our presentation and suggest possible further uses of the metaphors.

THE METAPHORS OF HUMAN THINKING

We describe thinking through five of its aspects which combined and separately catch highly important general properties that seem to be shared by human beings. Each aspect is described also by a metaphor meant to support the reader in keeping a clearer understanding useful in further

* The authors contributed equally to the paper.

studies and discussions. The five aspects of human thinking emphasized are habit, stream of thought, awareness, utterances, and knowing.

We have chosen to present these aspects of human thinking by quotations from James [20] and Naur [25-27]. Naur has carefully studied the 1377 pages of James's book *The Principles of Psychology* and through quotations, summaries and extended discussions illuminated James's work and to us made it more accessible. For readers who might not be aware of the continued importance of James's classical work in psychology, and who therefore might feel uncomfortable with our paper's building so directly on sources published more than hundred years ago, we quote the renowned cognitive psychologist Bernard Baars who in 1997 writes:

> 'Remarkably, the best source on the psychology of consciousness is still William James's elegant 'Principles of Psychology', first published in 1890. [...] James's thought must be understood in historical context, but the phenomena he describes so well have not changed one bit.', [2], p. 35.

For the purpose of improving our understanding of human thinking, we have not found any sources in psychology better suited than *The Principles of Psychology*.

The Eroded Landscape Metaphor of Habits

Every person is like a landscape eroded by water. By this metaphor we mean to indicate how a person's formation of habits leads to more efficient actions and less conscious effort, like a landscape through erosion adapts for a more efficient and smooth flow of water. Creeks and rivers will, depending on changes in water flow, find new ways or become arid and sand up, in the same way as a person's habits will adjust to new circumstances and, if unpracticed, vanish.

According to James the most important general property of the thinking and behavior of people is that each person is a bundle of habits. Building on James, Naur writes [27]:

> 'All our grasping of things around us that we see, hear, feel, that which we call perception, is entirely a question of the habits each of us has trained. In addition our locomotion, the way we move our arms and legs while moving around, is almost entirely habitual. In addition, our talking with each other, the way we grasp what others say to us and the way we move our tongue, lips, and other organs of speech while talking, all this has been trained as habits. All education is a matter of training habits.
>
> Any part of a human organism may be involved in a habit. In a certain sense every habit involves the entire person.'

Further, James discusses the possible physiological basis of habits which also sheds light on the nature of habits:

> 'Plasticity, then, in the wide sense of the word, means the possession of a structure weak enough to yield to an influence, but strong enough not to yield all at once. Each relatively stable phase of equilibrium in such a structure is marked by what we call a new set of habits. Organic matter, especially nervous tissue, seems endowed with a very extraordinary degree of plasticity of this sort; so that we may without hesitation lay down as our first proposition the following, that the phenomena of habit in living beings are due to the plasticity* of the organic materials of which their bodies are composed. *Note: In the sense above explained, which applies to inner structure as well as to outer form.' [20], vol. I, p. 105.

Human Thinking as a Stream of Thought

The metaphor of human thinking as a stream of thought is the result of James's own choice. He says [20], vol. I p. 239:

> 'Consciousness, then, does not appear to itself chopped up in bits. Such words as 'chain' or 'train' do not describe it fitly as it presents itself in the first instance. It is nothing jointed; it flows. A 'river' or a 'stream' are the metaphors by which it is most naturally described. *In talking of it hereafter, let us call it the stream of thought, of consciousness, or of subjective life.*'

Naur summarizes James's description of human thinking as stream of thought in this way [26], p. 85:

> 'In William James's Principles of Psychology the stream of thought denotes something happening in all of our wake moments, to wit our experience of thinking and feeling. The stream of thought is known to every one of us through introspection, that is through our turning the attention inward, towards the way we experience our thoughts and feelings. What we may register through introspection is merely a picture of rough outlines. The stream of thought changes incessantly and has a vast number of details, most of which are present only vaguely, far more than may be seized by introspection.
>
> The stream of thought happens independently of our desire. We may, when we so wish, more or less successfully think of something definite, but we cannot make the stream of thought cease, as experienced by every person suffering from insomnia.
>
> The stream of thought may be described as something that flows, an incessantly changing, complicated mixture of something that may be denoted explicitly as images, sounds and bodily impressions, with additional vague moods and feelings. As stressed by James we do not in the stream of thought experience sharply delimited parts or elements of any kind. At each moment our thought is occupied by something that is complicated, but that is experienced as a whole. These wholes James calls thought objects [Our remark: also called 'mental objects']. Within each thought object one may distinguish between something more at the center, that which is the subject of our attention, and something

that forms a fringe. […] [E]very thought object embraces feelings, including those of the personal well-being, moods and bodily presence.

In its continued changing the stream of thought alternates between substantive states of relative repose and transitive states of rapid change. During the transitive states the changes of the thought objects happen so rapidly that they cannot be seized by introspection.

In the experience of the stream of thought the present moment has a duration of a few seconds. As one thought object fades away by being replaced by another one, it is retained in the fringe of the coming one. Every sudden impression is always experienced as a whole with what was there immediately before it happened.'

Awareness as a Jumping Octopus

'The mental activity is like a jumping octopus in a pile of rags', says Naur [25] and continues to illustrate the dynamics of thinking:

> 'This metaphor is meant to indicate the way in which the state of consciousness at any moment has a field of central awareness, that part of the rag pile in which the body of the octopus is located. The arms of the octopus stretch out into other parts of the rag pile, those parts presenting themselves vaguely, as the fringes of the central field. […] The jumping about of the octopus indicates how the state of consciousness changes from one moment to the next.'

The rags of the pile may through focusing come to the field of central awareness. Here associations play a central role. On this Naur [26], p. 11, summarizes from James:

> 'One object of thought is replaced habitually by the next. We say then that the two thoughts are associated or that the next thought appears through its association to the first one. [...] [W]hat enters into the association of thoughts is not elementary 'ideas', but complicated *thought objects* which are experienced as wholes but each of which includes more central parts and a *fringe* of vague connections and *feelings*.'

Associations may happen by contiguity and by similarity. Association by contiguity is essentially a matter of habit formation. James [20], vol. I, p. 561 says:

> '[...] objects once experienced together tend to become associated in the imagination, so that when any one of them is thought of, the others are likely to be thought of also, in the same order of sequence or coexistence as before. [...] it expresses merely a phenomenon of mental habit, the most natural way of accounting for it is to conceive it as a result of the laws of habit in the nervous system.'

Association by similarity is:

> '[…] association between thought objects that have become connected in the thought merely by having the same abstract property in common, in other words by being similar in some respect.' [26], p. 12.

Association by similarity plays an important role in reasoning. Reasoning is concerned with solving problems, or answering questions, related to situations involving certain known things, having certain known properties, in which the person cannot reach the solution or the answer by direct association from the known properties. James explains how successful reasoning builds upon the person's noticing and attending to certain definite properties of the situation at hand, to wit such properties that point to a way of reaching the goal by direct association. James makes clear how reasoning in this sense is a decisive factor in human inventiveness and discovery, including that of scholars and scientists, see [27].

Utterances as Splashes over the Waves

'A person's utterances relate to the person's insights as the splashes over the waves to the rolling sea below', says Naur [25] and continues:

> 'This metaphor is meant to indicate the ephemeral character of our verbal utterances, their being formed, not as a copy of insight already in verbal form, but as a result of an activity of formulation taking place at the moment of utterance.'

The metaphor also emphasizes how utterances are vague and incomplete expressions of the complexity of a person's current mental object, in the same way as the splashes tell little about the sea below.

Human Knowing as a Site of Buildings

Human knowing is like a site of buildings in an incomplete state of construction, developed through maintenance and rebuilding. In Naur's [25] formulation:

> 'A person's insight is like a site of buildings in incomplete state of construction. This metaphor is meant to indicate the mixture of order and inconsistency characterizing any person's insight. These insights group themselves in many ways, the groups being mutually dependent by many degrees, some closely, some slightly. As an incomplete building may be employed as shelter, so the insights had by a person in any particular field may be useful even if restricted in scope. And as the unfinished buildings of a site may conform to no plan, so a person may go through life having incoherent insights.'

USING THE METAPHORS

Below we show how the metaphors can describe human-computer interaction phenomena known from research and commonly available user interfaces. For each metaphor, we describe examples that are coherent or in conflict with the metaphor, and an example where a notion commonly used in HCI with the aid of the metaphors appear to us as described simpler and clearer.

Habit in HCI

There is an abundance of examples of user interfaces that violate human habits. One example is adaptive menus, used for example in Microsoft Office 2000 [22]. Adaptive menus change the layout of the menu according to how often menu items are used, for example by removing or changing the position of items seldomly used. However, adaptive menus make it impossible to form habits in the selection of menu items [32], since their position may be different from when they were previously selected. A study by Somberg [34] showed the efficiency of constant position placement of menu items compared to menus that change based on use frequency. Somberg, however, did not explicitly link habit formation to the usefulness of constant placement of menu items. Note that the common practice of adding a fixed number of, say, recently used files or fonts to the bottom or top of a menu does not interfere with habit formation and may decrease time taken to select a menu item [33].

The discussion of consistency in user interfaces may be illuminated in terms of habit. In a classic paper on consistency [16], Grudin argues that focusing on consistency per se leads to a lack of focus on users and their tasks. In several examples he show how consistency can be interpreted in different ways and how different aspects of usability contradict each other in what some call consistent designs. From our point of view, Grudin's critique of the notion of consistency concerns the role of habit in the interface. With a focus on habits, the aim of consistency is to allow the habits that users develop to be transferable within or between systems they use. In addition, a system should also allow effective habits to be established in the first place, especially for often-used functions. Consistency between systems is not critical if interface elements or functions are not a habitual part of the users' repertoire of actions. Habitual association of words, however, might be useful for grouping or naming interface elements.

The central design issue with respect to consistency, and thus habit formation, is whether to utilize existing habits in the design of the system or create new ones. Grudin's [16] discussion of choosing effective keyboard layouts (e.g. QWERTY or DVORAK) is an example where it is essential for users to establish effective habits, rather than transferring real-world habits (such as associating letters in alphabetical order) to the interface. One reason why consistency is a problematic notion is that it obscures long-term usability—especially the efficiency gained by supporting inattentive, i.e. habitual, use. Perhaps designers in HCI more often should aim for establishing new, effective habits. Even the most radical changes of interfaces may be mastered if the interface is used often. An analogue of this is shown in Stratton's experiments with glasses that turned his visual field upside down [15]. When wearing the glasses constantly, in less than 7 days he had become habituated to viewing the world upside down and could walk, write, etc.

An example of a user interface that exploits that habit formation is not always wanted, is found in the evaluation version of the compression utility WinZip [37]. When WinZip is run, an initial screen with five buttons is shown. Three buttons allow the user to get access to license information, to a screen for registration, and to information about how to order. The last two buttons are of interest here. One button quits the utility; another lets the user proceed to the main screen of WinZip. To prevent users from going straight to the main screen, the designers of WinZip randomly interchange the position of the two buttons when the utility is run. Effectively, this prevents the user from establishing a habit of clicking the proceed button without noticing the license and ordering information on the initial screen.

Walker et al. [36] compare two different designs of a spoken language interface to email: (a) a mixed-initiative dialogue, where the users can flexibly control the dialogue, and (b) a system-initiative dialogue, where the system controls the dialogue. The results show that even though the mixed-initiative dialogue is more efficient, users prefer the system-initiative dialogue. A correlation analysis with user satisfaction as the dependent variable uncovers how:

> 'Users' preferences are not determined by efficiency per se, as has been commonly assumed. One interpretation of our results is that users are more attuned to qualitative aspects of the interaction.', [36], p. 587.

The number of automatic speech recognition rejects contributed the most to user satisfaction. Walker et al. suggest that the users' preference for the system-initiative dialogue arises from it being easier to learn and more predictable. This result was contrary to the authors' initial hypothesis. Evaluated from the aspect of habit formation especially the speech recognition rejects must be damaging. Even though the system-initiative dialogue requires a larger number of dialogue turns, this interface is preferred because it better supports habit formation.

Stream of Thought in HCI

A simple, yet effective, attempt to recreate part of the richness of the stream of thought when users return to resume interrupted work, is Raskin's design of the Cannon Cat [32]. When the Cannon Cat is started, the display immediately shows up as it was before work was suspended. Not only does this allow the user to immediately start thinking about the task at hand. It also provides help in remembering and recreating the stream of thought as it was when work was interrupted.

The fragility of the stream of thought is not well protected in many user interfaces. E-mail notifications, instant messengers, news on demand, automatic spelling and grammar corrections are useful at times, but may also disrupt concentrated work. Research on instant messengers, for example, has documented the harmful effects of interruptions on task completion time [11]. As a personal note, one of the authors of this paper has recently removed all notifications of arriving e-mails from his computer. Even the .5 cm $\times$.5 cm icon in the lower right corner of the screen that show the arrival of new e-mail could create an intense feeling of urge to check the e-mail—which would initially be in the fringe of the current object of attention, but eventually would lead to start of the e-mail program. This seemed especially to happen when that author was struggling with a difficult task. In general, we find that most user interfaces fail to support shifting between what we experience as two phases of work: concentrated working, where interruptions and distractions are detrimental, and explorative working, where a free flow of associations, inspirations, breaks, and even interruptions can be useful.

An example of the dynamics of thinking that is closely related to the stream of thought is found in information retrieval studies concerning changes in relevance judgments of documents. One study [12] showed that the order in which subjects viewed document descriptions influenced the subjects' perception of the relevance of those descriptions. While this effect in part may be due to the categorical rating scales used, a psychological explanation is also possible. When looking at document descriptions, the themes of the previous descriptions will be in the fringe of the subject's mental object. Those fringes will influence the perception of the task and the judgment of the current document description. Thus, different orderings of documents will give different relevance judgments. The study also describes how significant differences in relevance judgments can be found even between random orderings of the documents to be judged. Thus, relevance judgments seem to be dynamic in a sense closely related to the metaphor of the stream of thought.

Awareness in HCI

The metaphor of the octopus is well illuminated with studies of awareness presented at previous CHI conferences, e.g. [14,17]. Common to these studies is an aspiration to design for peripheral awareness, to design also for the fringes of the octopus so to speak. As an example consider Grudin's study [17] of how multiple monitors are used. Grudin found that among 18 users who used multiple monitors simultaneously, the multiple monitors were not used as additional space, but to partition the information used. Users would for example delegate secondary tasks such as debugging windows in a programming environment to the second monitor, and some users would have e-mail, news alerts, and instant messengers on the secondary monitor. Grudin's study is coherent with and supportive of the metaphor of awareness in two important ways. First, users employ the degree of attention they give information as a principle for dividing their work between monitors. Less important information is in the periphery of the eye and thereby to some extent in the fringes of the current mental object. This may reflect how subject introspectively realize that some information sources may in subtle ways distract us, but that they may be useful for creating fringes. Second, Grudin's work and other recent papers on awareness show opportunities for designing for peripheral attention and even in-attentive use of computers [35]. It is evident from the metaphor of the octopus that the fringe of mental objects form a large part of our thinking and this should be taken into account when designing.

The characteristics of awareness and the association of objects thought of with other objects are not unfamiliar descriptions of human thought in HCI. Vannevar Bush's vision of the Memex [5] may exemplify this:

> 'When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path.
>
> The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature. Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it.'

However, as pointed out by Wendy Hall at the Hypertext'01 Conference, links that take the user to web pages associated with the link description are fairly uncommon at the web [18]. In hypertext research, such links are called associative or referential links [9], as opposed to for example navigational or organizational links. According to Hall, less than 1% of links on the World Wide Web are associative: the rest are predominantly navigational links. On one side this suggests that Bush's warning has been taken seriously—human awareness and association are not directly modelled on the WWW. On the other side, we feel that the lack of associative links might suggest that designers have paid too little attention to awareness, associations, and how to craft links that use this fundamental trait of human thinking.

As an example of a notion in HCI that may become clearer from the metaphor of the octopus, we would like to briefly discuss information scent. Information scent refers to:

> '… the (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues, such as bibliographic citations, WWW links, or icons representing the sources' [30].

In HCI this notion has recently received much attention in relation to web design [7]. From our perspective, information scent is the ability of proximal cues to create in the mind of the user associations related to the content looked for. The degree to which WWW links or icons have 'information scent' is only a matter of the associations they create for individual users. In some studies of information scent, e.g. [31], an information scent score is developed. Subjects are given the top levels of a hierarchical link structure and the information scent score is the proportion of subjects who correctly identify that a certain link contains the answer to some task. Thus, subjects assess the links from the associations created in relation to the task. The second aspect of the definition of information scent—the cost of accessing information sources—is related to habit. We most often follow our habits in traversing information structures rather than pondering the cost of certain ways of navigation. That way, information scent is adequately described by the metaphors of awareness and habit.

Utterances in HCI

One consequence of the metaphor of utterances as splashes over the ocean is that we must expect users to describe the same objects and functions in an application program in a variety of ways. Furnas et al. [13] investigated the diversity in words used for describing commands and everyday objects. On the average, two participants described the same command or object by the same term with less than 20% probability. The most popular name was chosen only in 15-35% of the cases. Furnas et al.'s suggestion for relieving this problem is called the unlimited alias approach. Instead of using a fixed set of words for commands and functions, the unlimited alias approach lets users enter any term they want. If the term is not in the range of terms initially suggested by the designer of the system—which the data of Furnas et al. and the metaphor suggest it often will not be—the system may interactively suggest appropriate commands or object names. This approach is coherent with the metaphor and uses interactivity to clarify the intentions of the user.

Examples of user interfaces that do not respect the metaphor of utterances are plentiful. Many of these involve systems that try to predict, given a few utterances, the needs and wishes of the user—something that is unlikely to succeed given the ephemeral and incomplete nature of utterances. One example is the attempt of the Office Assistant in Microsoft Word to infer which kind of document the user is writing given one or two words.

We believe that the relation between queries made on the WWW and what users are looking for may be made easier understandable by use of the metaphor. Queries on the WWW are on the average 2.2 words long [21]. However, such short queries cannot possibly reflect all aspects of the pages users are looking for, nor can they reflect the myriads of interests, questions, etc. that may suddenly become the locus of attention when triggered by otherwise irrelevant web pages. In information retrieval, the difficulty in interpreting the intention (or information need) behind the queries has long been recognized as problematic, as have the difficulty of expressing one's information need in the first place [4]. Harter [19] has gone as far as to suggest that the information need is indeed our full mental constitution—which is impossible to express in a few words or queries. This is in accordance with the metaphor of utterances as splashes over the ocean and respects the complexity of mental objects, as described by the stream of thought and the octopus metaphors.

Human Knowing in HCI

One example that shows how effective it can be to respect the incomplete and developing character of human knowing, is found in object oriented programming, for example in the class libraries sometimes used to support development of user interfaces. Users of class libraries do not have to know the internal workings of the classes. Thus, they can program without having a complete understanding of the classes they use and gradually build up an understanding of how the class works, should that be necessary. The intuition from the metaphor would be that object oriented programming would give a faster and broader understanding of the program. A recent empirical study [10] treats differences of program comprehension during maintenance between 30 expert programmers of object oriented and procedural languages. The study suggests that the initial phase of program understanding is easier in OO programming languages because programmers gradually build their understanding from partial insights about a large part of the program:

> 'The OO programmers tended to use a strongly top-down approach to program understanding during the early parts of familiarization with the program, but used an increasingly bottom-up approach during the subsequent maintenance tasks. The procedural programmers used a more bottom-up orientation even during the early phase, and this bottom-up approach became even stronger during the maintenance tasks.' [10], p. 1.

However, the study also suggests that eventually both the OO and the procedural programmers built a systematic understanding of the program.

Examples where the metaphor of a person's knowing is not respected are easy to find. Systems that require a full understanding of the system before they may be used are cases in point. An example is described in Chen & Dhar's study [8] of an online library catalogue. They observe how 30 subjects take wrong actions in using the system, how they use wrong query terms, and how they use sub-optimal procedure for accomplishing tasks. The faulty actions arise from the subjects' misconceptions about the topic they are searching for, about the way the online catalogue works, and about the nature of the classification system used. Each subject displayed at least one misconception. First of all this shows that even for a common task like searching a library system, the subjects' knowing about the program was incomplete. Second, Chen & Dhar's results show that the design of the online catalogue violated the metaphor of the site of buildings in several ways. As one example, the system only recognizes official Library of Congress subject headings, which in essence requires the subjects to have a complete and precise understanding of how their problem relate to the official terms. The lack of support for cross-referencing and inferring correct headings worsen this.

Mental models have been extensively discussed in HCI. Consider as an example Norman's [29] description of the use of calculators. He argues that the use of calculators are characterized by users' incomplete understanding of the calculators, by the instability of the understanding, by superstitions about how calculators work, and by the lack of boundaries in the users' understanding of one calculator and another. These empirical observations by Norman are coherent with the ideas expressed by the metaphor of knowing. In summary, the OO programming example, the library catalogue, and the use of calculators show that users solve the actual tasks despite inconsistencies and incompleteness of their knowing. Conversely, systems that require a precise and complete understanding are awkward to use.

DISCUSSION

Readers who consider this description of human thinking to be mainly common sense may examine Table 1. We find it striking how essential descriptive terms in psychology of human thinking such as habit, thought, and knowing/knowledge are virtually absent from the CHI Conference Proceedings. Further, our examples have shown how an extended awareness of the five aspects of human thinking here emphasized can be useful in understanding important qualities of user interfaces and selected notions in HCI.

The aim of the paper was to describe HCI issues in the context of human thinking. We have not attempted to provide novel designs—readers with this interest should consult Raskin's work [32] for examples. More systematic exploration of the possibilities in design of using descriptions of the human thinking activity is desirable.

Word	CHI 2001	CHI 2000	CHI 99	CHI 98	CHI 97
Habit/ automati-city/auto-matization	0	0	0	0	0
Thought/ Thinking/ Think	2	2	0	1	0
Association	0	0 (1)	0	0	0
Awareness/ Aware	6 (1)	3 (1)	4 (1)	3	2
Utterance	1	0	0	0	0
Knowing/ Know/ Knowledge	3 (4)	2 (6)	2 (1)	0 (1)	1 (3)

Table 1—Number of papers in CHI Conference Proceedings 1997-2001 containing specified words in titles, keywords, or abstracts that describe human thinking (as found by searching ACM's Digital Library, September 2001). Numbers in parenthesis show the number of papers with only non-psychological uses of the word, as in '…little is known about…'.

For evaluation, one idea would be to develop from the metaphors a usability evaluation approach, similar to expert inspection techniques such as heuristic evaluation [28]. The metaphoric descriptions are psychologically recognizable and may be more inspiring to use and create more associations for the evaluators compared to e.g. heuristics or guidelines. Further, the metaphors may serve to uncover certain types of usability problems not found with traditional evaluation methods. Such problems might concern how well the interface supports habit development, the use of utterances in the interface, and the associations created by functions and descriptions of commands. However, to investigate the viability of this idea a series of experiments are needed.

The metaphors offer a high-level description of aspects of human thinking, whereas cognitive models commonly discussed in HCI, e.g. GOMS [6] or Interacting Cognitive Subsystems [3], focus on detailed descriptions of the operations and goals involved in solving tasks. Therefore, the metaphors may more conveniently create focus on human thinking, from early design ideas through evaluation to implementation and maintenance.

CONCLUSION

The human thinking activity was summarized through quotations from the work of William James and of Peter Naur. General properties of thinking activity known to all of us by introspection were emphasized through five metaphors. The metaphors catch psychological aspects of habit formation, stream of thought, awareness, utterances, and knowing. From commonly available user interfaces and from a selection of empirical studies, the utility of the metaphors was illustrated by showing their ability to clarify designs and notions in HCI. Since the metaphors address basic aspects of thinking, we suggest that the metaphors will be useful in design and evaluation of user interfaces. With the possible exception of awareness, these aspects of human thinking are virtually absent in recent years of the CHI Conference Proceedings.

ACKNOWLEDGEMENTS

[Removed for anonymous reviewing]

REFERENCES

1. Baars, B. J. *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge, 1988.
2. Baars, B. J. *In the Theater of Consciousness*, Oxford University Press, New York, 1997.
3. Barnard, P., May, J., Duke, D., & Duce, D. Systems, Interactors, and Macrotheory, *ACM Transactions on Computer-Human Interaction*, *7*, 2 (2000), 222-262.
4. Bates, M. The Design of Browsing and Berrypicking Techniques for the on-Line Search Interface, *Online Review*, *13*, 5 (1989), 407-431.
5. Bush, V. As We May Think, *The Atlantic Monthly*, July (1945).
6. Card, S., Moran, T., & Newell, A. *Human-Computer Interaction*, Erlbaum, Hillsdale NJ, 1983.
7. Card, S., Pirolli, P., Van Der Wege, M., Morrison, J. B., Reeder, R. W., Schradley, P. K., & Boshart, J. Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability, in *Proceedings of CHI 2001* (Seattle, WA, Mar 2001), 498-505.
8. Chen, H. & Dhar, V. User Misconceptions of Information Retrieval Systems, *International Journal of Man-Machine Studies*, *32* (1990), 673-692.
9. Conklin, J. Hypertext: An Introduction and Survey, *IEEE Computer*, September (1987), 17-41.
10. Corritore, C. L. & Wiedenbeck, S. An Exploratory Study of Program Comprehension Strategies of Procedural and Object-Oriented Programs, *International Journal of Human-Computer Studies*, *54* (2001), 1-23.
11. Cutrell, E., Czerwinski, M., & Horvitz, E. Notification, Disruption, and Memory: Effects of Messaging Interruptions on Memory and Performance, in *Proceedings of Interact 2001*.
12. Eisenberg, M. & Barry, C. Order Effects: A Study of the Possible Influence of Presentation Order on User Judgments of Document Relevance, *Journal of the American Society for Information Science*, *39*, 5 (1988), 293-300.
13. Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. The Vocabulary Problem in Human-System Communication, *Communications of the ACM*, *30*, 11 (1987), 964-971.
14. Greenberg, S. & Rouding, M. The Notification Collage: Posting Information to Public and Personal Displays, in *Proceedings of CHI 2001* (Seattle, WA, Mar 2001), ACM Press, 514-521.
15. Gregory, R. L. *Eye and Brain*, Princeton University Press, Princeton, NJ, 1997.
16. Grudin, J. The Case Against User Interface Consistency, *Communications of the ACM*, *32*, 10 (1989), 1164-1173.
17. Grudin, J. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use, in *Proceedings of CHI 2001* (Seattle, WA, Mar 2001), ACM Press, 458-465.
18. Hall, W. Mostly Linkless, in *Proceedings of Hypertext'01* (Århus, Denmark, Aug 2001), ACM Press, 3.
19. Harter, S. P. Psychological Relevance and Information Science, *Journal of the American Society for Information Science*, *43* (1992), 602-615.
20. James, W. *Principles of Psychology*, Henry Holt & Co., 1890.
21. Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. Real Life Information Retrieval: A Study of User Queries on the Web, *SIGIR Forum*, *32*, 1 (1998), 5-17.
22. Microsoft Office 2000, www.microsoft.com.
23. Naur P., Human knowing, language, and discrete structures, 1988, In *Computing: a Human Activity*, ACM Press/Addison Wesley, 1992.
24. Naur, P. *Knowing and the Mystique of Logic and Rules*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1995.
25. Naur, P. CHI and Human Thinking, in *Proceedings of NordiCHI 2000* (Stockholm, Oct 2000), Available from www.naur.com.
26. Naur, P. *Anti-philosophical Dictionary*, naur.com publishing, 2001.
27. Naur, P. *Psykologi i videnskabelig rekonstruktion (*Eng. *Psychology in Scientific-Scholarly Reconstruction),* 2001, under publication.
28. Nielsen, J. *Usability Engineering*, Academic Press, San Diego CA, 1993.
29. Norman D., Some Observations on Mental Models, in Gentner, D. & Stevens, A.L. (Eds.), *Mental Models*, Erlbaum, Hillsdale NJ, 1983, 7-14.
30. Pirolli, P. & Card, S. Information Foraging, *Psychological Review*, *106*, 4 (1999), 643-675.
31. Pirolli, P., Card, S., & Van Der Wege, M. The Effect of Information Scent on Searching Information Visualizations of Large Tree Structures, in *Proceedings of AVI'2000* (Palermo, Italy, ACM Press), 161-172.
32. Raskin, J. *The Humane Interface: New Directions for Designing Interactive Systems*, Addison-Wesley, Reading MA, 2000.
33. Sears, A. & Shneiderman, B. Split Menus: Effectively Using Selection Frequency to Organize Menus, *ACM Transactions on Computer-Human Interaction*, *1*, 1 (1994), 27-51.
34. Somberg, B. L. A Comparison of Rule-Based and Positionally Constant Arrangements of Computer Menu Items, in *Proceedings of CHI+GI'87* (Toronto, Canada, 1987), ACM Press, 255-260.
35. Strøm, G. *Inattentive Use of Electronic Equipment*, PhD Thesis, Department of Computer Science, University of Copenhagen (1996).
36. Walker, M., Fromer, J., Di Fabrizzio, G., Mestel, C., & Hindle, D. What can I say?: evaluating a spoken language interface to Email, in *Proceedings of CHI'98* (Los Angeles, CA, Apr 1998), 582-589.
37. Winzip, Version 7, www.winzip.com.