DISPLAYING MEDICAL LABORATORY REPORTS ON SMALL SCREENS

by

Heidi Lam

B.A.Sc., Simon Fraser University, 2001

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in the School

of

Engineering Science

©Heidi Lam 2004

SIMON FRASER UNIVERSITY

April 2004

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without permission of the author

Approval

Name: Heidi Lam

Degree: Master of Applied Science

Title of Thesis: Displaying Medical Laboratory Reports on Small Screens

Examining Committee:

Dr. Rodney Vaughan
Professor
Chair

Dr. John Dill
Professor
Senior Supervisor

Dr. Stella Atkins
Professor
Supervisor

Dr. Arthur Kirkpatrick
Professor
Supervisor

Dr. Andrew Rawicz
Professor
External Examiner

Date Approved:

SIMON FRASER UNIVERSITY

Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada

Abstract

Medical laboratory testing is an important tool for physicians in disease screening, diagnosis and progress monitoring. Currently, laboratory reports are printed on paper. Despite their popularity, paper-based reports are slow, access-limited and difficult to share. In view of these problems, this thesis investigates methods of delivering laboratory results to personal digital assistants (PDAs) electronically. The prototype was designed following the user centered design process, and based on investigations on the characteristics of the target users, the laboratory report interpretation task, the characteristics of the PDAs, and the nature of the laboratory data. The main features of the design include an anatomical overview for rapid patient status assessment, the use of a hierarchical + elision presentation technique for better diagnostic support, and the provision of disease profiles for better patient management. In addition, various visualization techniques were used to minimize limitations imposed by the small screen size, limited CPU processing power and slow and intermittent wireless connectivity. The prototype was evaluated in an observational user study with eight physicians and received very positive feedback. The hierarchical + elision technique was further studied to quantify its potential costs and benefits as a presentation technique for lists on small screens. The comparative user study indicated that in general, the hierarchical + elision technique can provide the same potential benefits and costs when compared to the conventional linear technique, and is well suited for displaying laboratory results.

Acknowledgments

I would like to express my deepest gratitude to my senior supervisor, Dr. John Dill, and my committee members, Dr. Stella Atkins and Dr. Arthur Kirkpatrick, for their invaluable support, advice, guidance and time given to me in the preparation and development of the current thesis.

I also wish to thank my participants for their help with my research.

Table of Contents

List of Tables

List of Figures

1 Introduction and Motivation

Medical laboratory testing is an important tool for health care professionals in patient status assessment, diagnosis, and progress monitoring. Traditionally, test results are reported in print, listing the test names, the patient's test results, and a range of nominal values. This method of result delivery poses two serious problems for the caregivers. First, how do they interpret the results to plan the next course of action for the patient? Second, how do they access and share these results in a timely manner?

1.1 Information Visualization: To make those numbers meaningful

Displaying test results effectively is not a trivial exercise. A single numerical test result in itself is not enough to support decision making for the caregivers. Other important reference information required includes test-specific and patient-specific information. Test-specific information includes reference range, date of sampling, diagnostic guidelines provided by the laboratory, and clinical information provided by the primary physician, while patient-specific information includes age, gender, height, and weight. In short, a typical laboratory report contains a large amount of data, and if these data are presented in a dull and illogical manner, there is a danger that important information will remain unnoticed (Mayer et al., 1998). In fact, studies have shown that even in prestigious medical institutions, available laboratory data fail to elicit an appropriate clinical response in 30% of cases (Altshuler, 1994), and 9% – 31% of laboratory errors can be attributed to inappropriate interpretation and utilization of laboratory results (Bonini et al., 2002). Information displays that fail to consider the nature of the laboratory data and human perceptual limitations may, at least partially, be responsible for this problem (Mayer et al., 1998; Wright et al., 1998).

1.2 A network application: To make those results available

The second problem encountered by caregivers with the current paper-based reports is delivery delays and limited access. In fact, the lack of immediate notification and/or clinical utilization of a critical value can have an effect on outcome as negative as a wrong result (Bonini et al., 2002). In addition, as patient care moves towards the multi-team and multi-disciplinary approach, information synchronization and coordination have become difficult and important tasks. With diverse geographic locations and computer technologies, providing access to the patient's record to all caregivers involved is not a trivial problem. When the results are not accessible, physicians tend to reorder the tests, leading to unnecessary health care spending and patient inconvenience. Inappropriate test use has been showed to be a severe problem where in some cases is over 90% (Bandolier, 2004). The financial impact is severe. In the United Kingdom, £1.6 billion is spent on costs of laboratories and laboratory testing. Even if a small percentage of inappropriate test spending is due to lack of information sharing, the potential saving in cost is still tremendous.

1.3 Thesis Outline

The accessibility problem can be partially addressed by electronic result delivery. Currently, there are desktop, web-based report delivery systems in use, but the requirement of a desktop computer limits the availability of the test results, especially when physicians visit different examining rooms, wards, and even hospitals regularly in their practice. Alternatively, results can be delivered to small mobile devices to better reflect the mobile nature of physicians' work style. In that case, the information visualization problem described earlier becomes even more difficult when the results are to be displayed on a small screen. Typically, mobile devices are severely limited in screen size (about 5cm x 5cm) and display capability (colour and resolution) when compared to their desktop counterparts.

The goal of this thesis is to investigate the information visualization issues surrounding the use of small screens in reporting numerical laboratory test results.

In order to design a system that is appropriate for the target users and the intended task, a study on various aspects of the target users and the laboratory results retrieval and interpretation task was conducted. Findings are categorized in four sections: User, Environment, Task, and Presentation of laboratory results. Based on these findings and research on human perception, a series of prototypes were designed following the user-centered iterative design process evaluated using observational user studies. Three detailed user studies were designed to further study specific aspects of the final design, from which the study of hierarchical versus linear displays was chosen as the subject of a comparative user study. The thesis concludes with a list of future directions, both for the prototype as a commercial product, and for the interesting questions arising from the comparative user study.

2 Users

Understanding the target users is the first, and most important step in interface design. This chapter describes the research conducted on the various characteristics of the target users in order to construct a user profile. The areas considered include medical cognition, user motivation, task and system knowledge, user job and task, and their physical characteristics. Information in these areas was gathered from literature research, a questionnaire given to physicians from various medical disciplines (Appendix A), and personal clinical experience as a medical student for three years.

2.1 Target Users

The target users of this laboratory results reporting system are physicians. In the current North American health care system, physicians act as the main decision makers and coordinators in patient management. They are therefore the main consumers of information presented in laboratory reports. Other health care professionals, including nurses, therapists and receptionists may consult and transcribe laboratory data, but they generally do not use them for decisions in patient management.

2.2 Cognition

Understanding the cognitive model in medical decision making and diagnostic reasoning is essential in designing an effective system for physicians, since laboratory results are usually used as part of clinical findings available to physicians for decision making. This section explores a number of aspects of medical decision-making and problem solving—general models, expert-novice differences, and clinical reasoning based on laboratory data.

2.2.1 Models of Medical Decision Making and Problem Solving

According to Patel et al. (1994), medical decision-making can be divided into two broad categories:

1. ***Fault Model***. Similar to diagnostic troubleshooting in electronics, physicians solve problems using a coherent knowledge structure built on their biomedical knowledge and clinical experience. The primary goal is to find a system fault or perturbation.

2. ***Heuristic Classification Model***. Instead of finding faults in the system, this model suggests that the goal of clinical reasoning is to classify a cluster of patient findings as belonging to a specific disease category. Clinical experience is therefore extremely important, as reasoning is guided by exemplars and analogy. Functional understanding of the system is used to resolve anomalies, and in therapy planning.

Patel et al. (1994) believe that diagnostic reasoning is one of heuristic classification, since basic science knowledge cannot be easily integrated in clinical contexts. In fact, they found experts only use detailed biomedical knowledge when they are solving difficult problems, or problems outside of their specializations, but not when solving routine clinical problems in their own domain of expertise.

Werner (1995) listed four strategies in medical decision-making and problem solving. All these strategies can be considered as belonging to the heuristic classification model.

1 ***Pattern Recognition***. Using this strategy, medical diagnosis is based on a single, highly probable hypothesis generated from early and fragmentary evidences collected from the initial history and physical examination. This strategy can be effective since most clinical cases are simple and straightforward, and most signals of real value are striking. Decisions made under this strategy are based on certainty of the working hypothesis, and the principle decisional task is to rule in the working hypothesis.

   While this strategy is effective most of the time, clinicians may attempt to validate the working hypothesis despite conflicting clinical evidences. In fact, some physicians may ignore clinical information that does not agree with their working hypothesis.

2 ***Hypothetico-Deductive Process***. Instead of forming and working on a single hypothesis, clinicians may work on a larger number of hypotheses, typically three. These hypotheses may be very specific, or very general. Each addition piece of information gathered in the process will be used to validate or invalidate the hypotheses. Other than likelihood,

physicians also consider the hypothesized problem's treatability and severity. Thus, under this strategy, the decisions are made under risk, with the principle decisional task being both ruling hypotheses in and out.

3 ***Arborization***. Arborization is a more formal decision-making strategy, in which physicians sequentially analyze pertinent data in a systematic, preset process. Based on available evidence, the physician selects one of the multiple available paths until a correct hypothesis is arrived. Decisions are still made under risk, since at each cross-point, the decision made may not be correct based on partial evidence. Similar to the hypothetic deductive process, the principle decisional task is to rule in and out of hypothesis. But for the strategy of arborization, decisions are made based solely on available evidence without any consideration for likelihood, treatability and severity.

4 ***Exhaustion***. Using this strategy, the physicians would accept all possible hypotheses without considering their probability. This is similar to screening of a disease, where tests are ordered indiscriminately to detect a potential problem. Exhaustion is decision made under total uncertainty, since any outcome is possible. The main decisional task is to collect and interpret all relevant medical facts without formulating a working hypothesis.

2.2.2 Expert-Novice Differences in Clinical Reasoning

The pattern in diagnostic reasoning among subjects of various degree of expertise differs. The different levels of medical expertise can be defined as follows (Patel et al., 1994).

- ***Novice***. An individual who has begun to acquire the prerequisite knowledge assumed in the domain, such as a medical student.

- ***Expert***. An individual with specialized knowledge of the domain, for example, an experienced cardiologist in the domain of cardiology.

- ***Subexpert***. An individual with generic knowledge, but not specialized knowledge in the domain. An example would be an endocrinologist in the area of cardiology.

In the case of novices, their level of expertise can be further classified as early, intermediate and advanced. A study by Arocha et al. (1993) shows that early novices (second-year medical students) tend to generate and retain their initial hypotheses corresponding to the most common disease, and maintain these hypotheses throughout the case despite conflicting evidences. Intermediate (third-year students) novices generate several hypotheses in parallel, but without resolving or eliminating them. In other words, those hypotheses are used to account for different aspects of the clinical data. Advanced (forth-year students) novices also generate multiple, but competing hypotheses, since they are used to account for the same set of findings. Also, advanced novices tend to evaluate these hypotheses against the clinical findings to rule out hypotheses.

Patel et al. (1994) found that experts generate their hypotheses fairly early in the process, and these hypotheses tend to be high-level. With subsequent data, instead of generating more hypotheses, the existing hypotheses are refined and evaluated. In other words, with an increasing amount of data, experts tend to decrease their uncertainty. Since the experts generate hypotheses based on clinical findings, their reasoning pattern is defined as forward reasoning. In difficult cases where data are suggestive of multiple competing diagnostic possibilities, experts do use backwards reasoning to tie loose ends.

In contrast, subexperts generate low-level hypotheses and continue to do so when new findings are revealed. Subexperts are therefore less likely to evaluate their hypotheses and thus unable to rule out hypotheses. In other words, with an increasing amount of data, subexperts tend to increase, rather than decrease, uncertainty. Since subexperts reason from hypotheses to data, their reasoning pattern is defined as backward reasoning.

Despite its effectiveness, pure forward reasoning is only successful in situations where knowledge of a problem can result in a chain of inferences from the initial problem statement to the problem solution. This requires domain knowledge, since there are no built-in checks for the legitimacy of the inferences. In contrast, backward reasoning is slower and may make heavy demands on working memory, because the clinician needs to keep track of goals and hypotheses. It is most likely to be used when domain knowledge is inadequate, as in the case of subexperts, and in difficult cases for experts.

Expert reasoning is believed to be schema-driven (Koedinger & Anderson, 1990; Hunt 1989). A schema is assumed to be a structure that represents generic knowledge and is instantiated with specific new information in a given situation (Brewer, 1987). Since schemas guide physicians to key elements and serve to filter out irrelevant information, effective schema formation is essential in generating good decisions. One impediment to schema formation is short-term memory limitations. In medical reasoning, the use of schema is dynamic. Based on their repository of schema in their long-term memory, physicians tend to construct a schema for each patient in their working memory. Experts are highly skilled at developing memory structures for maintaining intermediate results, such as unaccounted for, but significant findings (Ericcsson & Staszewski, 1989). This memory encoding and retrieval ability is critical for maintaining and evaluating hypotheses in working memory, especially when clinical findings tend to be presented sequentially, rather than simultaneously. Norman et al. (1989) found that experts recall more data than novices. Even though both experts and novices recall more critical and abnormal findings than the non-critical ones, experts tend to recall more non-critical data than novices, probably since the experts devote more resources to refine and evaluate their hypotheses, and have more structured schemata to aid information storage.

2.2.3 Clinical Reasoning based on Laboratory Tests

Patel et al. (1997) reported that despite ordering the same number of tests, physicians with different levels of expertise use the tests differently. For experienced doctors, the reasoning has two components. In the main component, every ordered test is linked to the main hypothesis while accounting for alternative hypotheses within the framework. The second, smaller component is to plan therapy.

For inexperienced doctors (e.g., interns), the tests ordered are not always related to the diagnosis structurally, despite having a good knowledge of the functions of the tests. Interns show a lack of relevance to the clinical context of the patient, and use a generic exploratory strategy to search for relevant cues to refine or elaborate their hypotheses. They are, however, able to relate the results of the tests to the diagnosis. In other words, even though the test requests may not be schema-driven, the diagnosis does provide a context for test interpretation.

For medical students, their approach to ordering tests tends to be unstructured. In other words, they tend to order tests without reasons. Despite understanding the underlying function of the tests, they do not have the clinical context to correctly interpret the results.

2.2.4 Conclusions from Cognition

Medical problems are complex problems, and physicians tend to be heuristic rather than algorithmic in their approaches. In addition, their reasoning approaches differ depending on their level of expertise and their domain knowledge. In order to facilitate effective problem solving for all users, the software system needs to be flexible enough to accommodate these differences. For experts, laboratory tests are used to refine and evaluate their hypotheses and to plan therapy. Since each test is ordered for a reason, experts only need to be shown the test result, which will be interpreted in the framework of the hypotheses. In this case, results should be displayed in a simple and straightforward manner. An example would be to display them in an alphabetically sorted list. For subexperts and novices, since their reasoning may not be structured and coherent, laboratory tests may be used to search for clues and thus generate more hypotheses, rather than to evaluate existing hypotheses. In that case, it would be helpful to structure the tests to facilitate a more structured reasoning. An example would be to group the tests according to anatomical or physiological systems, like endocrinology, cardiology, and so on. This organization may be frustrating if the user is looking for a specific test, since there may be cases when a test can be potentially grouped under more than one system. Also, some tests may be difficult to categorize. Examples include glucose level, and the electrolytes, which are systemic. In short, the users should be able to choose between these display techniques.

2.3 Motivation and Attitude towards System

Certain psychological characteristics contribute to the performance and expectations of users. In the task of using laboratory results, physicians' motivation and attitude towards computers in general, and electronic result-reporting systems in particular influence their acceptance of, and adaptation to the new laboratory result reporting system.

2.3.1 General Attitude towards Computer Systems

The main concern about the target user group is their attitude toward computer use in their practice. Many reports attribute the lack of computer acceptance in health-care settings to physicians' work, attitudes, interests, and enthusiasms, but not to technical problems (Appleby, 1997; Brooks, 1999; Shumway, 1990; Young, 1984). From personal observations, physicians tend to be conservative in selecting and using new tools as, over more than a century, medical practice has been established and refined to the form adapted by most practicing physicians. This may be partly due to the conservative nature of medicine, and partly due to the tight schedules in many clinical practices where patients are often scheduled at fifteen-minute intervals with little or no time between visits (Shumway, 1990). It is therefore difficult for them to perceive enough potential benefits in adding computers to their familiar work processes to justify the heavy initial time investment in training and habit-reforming with their heavy work-loads. In fact, a study of physicians' attitude towards the use of computers in accessing drug information stated that the introduction of computerized information databases in clinical practices may not fit in with the work-style the physicians have adopted (Shumway, 1990; Garrard, 2000), and physicians believed the use of computer would decrease their productivity (Appleby, 1997).

Due to their demanding schedules, and the vast amount of new and existing medical knowledge required of physicians, they have very little time to explore areas that are not directly related to medicine. This has two consequences. First, physicians tend to be less knowledgeable about the benefits that could be obtained through the use of computerized information resources, and consequently, are more skeptical concerning the role of computer information systems in reducing the costs and improving the quality of health care (Shumway, 1990). Second, they tend to be less familiar with popular computer applications and devices, including the ubiquitous web browsers and PDAs.

However, the situation is changing. Market data suggests that the number of physicians using the Internet is growing. In the United States, the percentage of physicians using the Internet was reported to be 37 to 63%, with the daily usage being 57% (eMarketer, 2001). The situation of PDA use among physician is similar. Harris Interactive (2001) reported that in 1999, 15% of physician reported using PDAs, while in 2001, the figure increased to 26%. However, many reports have stated that the frequency of computer use by physicians is substantially lower than that the rest of society (Appleby, 1997; Brooks, 1999; Shumway, 1990).

This unfamiliarity may lead to something called "computerphobia". Physicians are trained to be authority figures. It is necessary for physicians to instill confidence in patients, patients' families, colleagues, and other health-care professionals. When being confronted with unfamiliar computer systems, the physicians may be unwilling to experiment to avoid displaying signs of "incompetence". In fact, the developers of InfoClique, a web-based patient information delivery system, listed "Beware of computerphobia" as one of the lessons learned (Brooks, 1999).

A related issue is physicians' perception of the role of computers in health-care. With the rapid development of expert systems, and the media's portray of a computer-dominated world, physicians may see computers as threats to their expertise, and which will reduce the human contact in health care. Nonetheless, Shumway (1990) showed that physicians believed themselves to be receptive to computer technology. Shumway (1990) conducted a self-reporting survey with physicians and found that physicians did not believe computer applications would affect their traditional role in health-care. They also reported the use of computer applications in their practices would not impede their individual achievement and recognition, reduce the need for medical manpower, alienate them from patients, result in reliance on "cook-book" medicine, depersonalize medical practice, or create liability problems. In fact, the study reported that physicians believed medical decision-making skills would be enhanced with the use of computers.

2.3.2 Motivation in Using an Electronic System

Due to their demanding schedules, it is unlikely that physicians will voluntarily explore and learn to use a new system without substantial motivation and perceived benefits. Depending on the hospital's and the laboratory's policies, physicians may be forced to use the new system if it completely replaces the old one. Even in those cases, it is known that physicians may ask the receptionists to print out the output of the system, instead of directly using it (Appleby, 1997). However, the advantage of enhanced accessibility of an electronic system may provide a strong motivation for physicians despite the perceived learning required.

2.3.3 Conclusions from Motivation and Attitude

Clinicians tend to be discretionary users whose computer usage is mostly by choice. They need to feel immediately that the system will not take too long to learn, will not impede their busy schedules, and will offer real benefits. Also, due to their fear of computer systems in general, the system should be more tool-like, be consistent and predictable, to give the users an immediate sense of control and accomplishment. The design priority for the system should be ease of learning.

2.4 User Knowledge and Experience with Existing Systems

Users' knowledge and experience with the task of retrieving laboratory results, and with existing result reporting systems tend to influence their expectations and eventual acceptance of the new system. Also, understanding their prior experiences with computer systems will provide a rough assessment of their computer skills, and interaction expectations.

2.4.1 Task Experience

Laboratory tests are routine clinical findings that are important in medical decision-making. All physicians are familiar with the purpose and function of most laboratory tests, but tend to use them differently depending on their level of expertise in the domain of the clinical case (§2.2.3). As for the task of test result retrievals, some physicians may have prior experience with existing electronic reporting systems where they need to locate the results themselves. However, most physicians are currently using the paper report form, which are received and filed by supporting staffs into individual patient charts. Another route of retrieval in the hospital setting is through direct call-in to the in-hospital laboratories. This is mostly frequently performed by medical staffs instead of attending physicians.

2.4.2 System Experience

With the prevalence of the Internet, most users would have some exposure to the desktop web-style interactions. Increasing numbers of physicians are using PDAs for medical references, especially for drug information. Experience with handheld devices will affect their expectations in a laboratory result reporting application.

2.4.3 Use of Other Systems

Traditionally, laboratory results are delivered as a paper report. This mode of information presentation will also affect the physicians' acceptance of and opinion about the new system, especially when they are familiar with the layout of the document (Nygren et al., 1992). Some users may have prior experience with existing electronic reporting systems.

2.4.4 Conclusions from User's Knowledge and Experience

There is no doubt that physicians are very familiar with laboratory tests, and the use of test results as part of their decision making tools. Their existing experience with the paper reports will very likely influence their acceptance of their electronic counterpart, since they are likely to be very familiar with the layout of the paper report. Also, in most cases, attending physicians do not need to retrieve the test results. In other words, the results are available to them when required. Using electronic systems will therefore impose an extra step in their busy schedule, since the physicians themselves will then be required to perform the task of retrieval. This imposes a requirement for the electronic system to allow simple and effective result retrievals.

2.5 User's Job and Tasks

The nature of the physician's job and the tasks will influence the design of the system. This section explores the impact of frequency of use, primary training, and system use.

2.5.1 Frequency of Use and Task Importance

Since laboratory tests are extremely important in many aspects of patient care, the frequency of use will be high, at least once per day. In other words, clinicians will have many learning or exploring opportunities with the system, and considerable incentive for learning. However, it has been found that some clinicians may pass the task to other medical staffs (Brooks, 1999). Also, the turnover rates of physicians in hospital settings can be very high, especially for teaching hospitals. New students, interns, and visiting doctors join and leave the clinics frequently. Their stays at the hospital tend to be short, from a few months to a few years. Also, in-hospital physicians may take temporary leave to visit other medical and research centres. In that case, despite the frequent use of the system, one cannot safely assume users will remember system-specific interaction techniques and commands.

2.5.2 Primary Training

Due to the independent and busy schedules of most physicians, formal training is difficult to schedule. It is therefore essential that the interface to simple and memorable. In addition, a printed manual should be available to physicians for at-home explorations of the system and self-studies, and as a general reference (Brooks, 1999).

2.5.3 System Use

Traditional paper based laboratory reports will likely remain in existence along with the new system. Use of the electronic system is therefore optional and voluntary, instead of mandatory. However, the new system will provide the benefits of convenience and speed of result retrieval, since the results will be displayed as soon as they are available from the laboratories, and related information will be accessible on the same display.

2.5.4 Conclusions from User Job and Tasks

Due to the importance of task and the relatively high frequency of use, the user may invest in time to learn to use the system. In other words, efficiency may be more important than ease of learning and recall. However, considering physicians' demanding schedules, the potentially high turnover rate and voluntary system use, ease of learning and recalling cannot be severely sacrificed for efficiency.

2.6 Physical Characteristics

It can be assumed that the physical characteristics of physicians are the same as in the general population. Consideration regarding colour blindness and handedness should be the same.

2.7 Conclusions from Users

The following implications may be drawn from the user profile constructed based on Mayhew's analysis (Mayhew, 1999):

1. ***Low motivation, discretionary use***. The design goal should be ease of learning.

2. ***Naïve or inexperienced users of an interactive system***. The system should be consistent, predictable, simple to understand and operate, and tool-like, rather than human-like, in order to give the user an immediate sense of control, mastery, and achievement. This will reduce frustration, fear, and stress, and increase motivation, thus allowing for better performance.

3. ***High task, low system experiences***. The system should have a minimal semantic prompting but a lot of syntactic prompting and instructions.

3 Environment

Interfaces are usually limited by the target display device. In this application, the target device is a personal digit assistant (PDA). A simple comparison between two kinds of PDAs and a typical desktop device highlights the two most important design constraints covered in this chapter: small display area and low processing power (Table 1).

	Desktop	Palm[1]	Pocket PC[2]
Screen Size (W x H cm)	36 x 27 [3]	6 x 6	6 x 8
Resolution (W x H pixels)	1600 x 1200	160 x 160	320 x 320
CPU Speed (MHz)	2 000 [4]	33	400
RAM (MB)	512 [4]	8	64
ROM (MB)	40 000	2	48

Table 1. Comparison between desktop, Palm and Pocket PC PDA in terms of screen size, resolution, CPU speed, RAM and ROM.

In addition, considerations are given to the specific interaction styles and input mechanisms for small mobile devices. The chapter then focuses on the issues involved in displaying data on small screens. A section that outlines techniques in transforming data display from a larger display area to smaller screens is also included in the chapter to provide further insights in issues concerning small screens data displays.

---

1 Values based on Palm m103

2 Values based on iPAQ H3970

3 based on a SONY 19-inch monitor

4 Intel Pentium IV

3.1 General Design Issues in Mobile Devices

There is a large amount of literature investigating the unique design issues in mobile devices with small screens (e.g., Kamba et al., 1996; Watters et al., 2002).

3.1.1 Small screen

This is the most prominent user interface challenge in designing for small devices. With 1/30 of the screen size and 1/10 of the resolution (Table 1), transforming information displayed on the desktop to the handheld is evidently difficult. In fact, various studies have shown that performance on small screens is less effective than on their desktop counterparts (Kamba, 1996; Jones, 1999), especially when the task is complex (Watters et al., 2002).

3.1.2 Limited input devices

Text input is difficult in mobile devices because most devices do not conveniently support conventional keyboard entry. Many PDAs do provide the option of a foldable keyboard. However, the effectiveness of this input device is controversial. For the rest of the users, input is done by stylus and soft-keys.

3.1.3 Limited CPU and battery life

Because of the size of the devices and the battery life, processing information on the devices is very expensive. As a result, only essential operations should be performed on the devises.

3.1.4 Limited bandwidth and Intermittent Connection

This is a special consideration for wireless applications. Due to the low bandwidth and uncertain connectivity of mobile devices, data transfer can be slow and sporadic.

3.2 Interactive Styles In Mobile Devices

Most small device studies have concentrated on general Internet interactions rather than on specific applications. Nonetheless, there are valuable lessons in these studies due to the similarity in user tasks and behaviours.

Studies have shown that mobile Internet interactions tend to be goal-directed (Jones, 1999). In other words, the mobile device will not be used primary for recreational, undirected browsing, but for extracting particular bits of information relevant to the current task. Usually, this process begins somewhere near the correct answer, and involves some amount of navigation to home in on the information target. Physicians' interactions with the result reporting system are also likely to be goal-directed. The search for a particular patient's report will be very specific, but the specificity in searching for individual tests within the report may depend on the physician's domain expertise.

It is widely agreed that average task times in small devices are significantly less than that in desktop (PalmSource, 2002; Jones, 1999). The Palm User Interface guidelines suggest basic sessions should be less than five minutes (PalmSource, 2002). Users are therefore less likely to tolerate long delays in data transmission and processing. This behaviour is consistent with the time-pressured environment of physicians.

Methods to address these concerns includes provision of direct searching functions, and display a list of user goals on a single page.

3.3 Displaying Data on Mobile Devices

Studies have investigated the best approach to display various data types on small screen (e.g., Buyukkokten et al., 2001 for texts; Watters et al., 2002 for lists and tables). The two data types of interest are tables and lists.

3.3.1 Tables

In current paper reports, numerical and quantitative laboratory results are displayed as tables. Watters et al. (2002) found that providing consistent context information improved the performance of users when looking up information from large tables on small screens. The same study also found that taking up part of the space of a small screen for search is not indicated for most situations.

3.3.2 Lists

An alternative for displaying laboratory results is lists. With smaller screens, it becomes impossible to view the whole list on one single screen. Maintaining context in the middle of a long list is therefore very important. One way to provide context information is to include a header and to provide the relative position of the list on each screen. In terms of displaying the list itself, some approaches include breaking the list up into pages, with "Back" and "Next" links for navigation, with or without a search function, and scrolling.

3.4 Transforming Data from Desktops to Mobile Devices

In order to display more information on these mobile devices than their screens can accommodate, Watters et al. (2002) grouped the various approaches into four main categories:

1. ***Dump-and-scroll***: This is the simplest method. Data is displayed on the small screen the same way as their desktop counterparts, with horizontal and vertical scrolling. Scrolling has been identified as one of the factors that affects reading rate. Line width has been shown to effect performance more than the number of lines in text comprehension (Duchnicky & Kwahk, 1983).

2. ***Tailor-made***: Content of websites are modified to be viewed on small screens, either by design, or with specific description languages such as WML WAP, 98, or CHTML by the 1998 W3Consortium .

3. ***Chunking***: This approach uses transformational rules to break up the data based on the data type, so that the resultant display fits the target device.

4. ***Zooming***: The zooming techniques allow viewing of large data area by the use of the focus + context visualization technique, thus allowing the user to designate and view an area of interest in detail, while the rest of the screen provides a low resolution context and serves as the background. Examples of such zooming technique include Table Lens (Rao & Card, 1995) and accordion summarization (Buyukkokten et al., 2001).

These techniques assume the Internet being the data source, and the nature of the data is unknown. In our situation, the data to be displayed on the small device is known beforehand, so a more specific and pre-defined interface can be designed. However, these transformation techniques are still very useful in the initial design stage, since in a sense, the design challenge is essentially the same—to fit data ideally displayed on large screens onto small ones.

3.5 Conclusions From Environment

It is evident that with the differences in device capabilities, it is unreasonable to directly copy existing laboratory desktop interfaces onto small devices. In fact, reviews comparing the 3Com PalmPilot with Windows CE devices often claim that Windows user interface style created for the desktop does not work well on palm-size devices (Lasky et al., 1998). Also, users do not necessarily expect small devices to act like "regular" computers. In order to customize the small screen interface, an understanding of the task is essential.

4 Tasks

This chapter describes the uses of laboratory results, and explores the task of both result retrieval and result interpretation in a mobile, small screen scenario to provide context for the design prototype. A desktop scenario is also outlined to highlight special considerations in the mobile small screen case.

4.1 Goals

In general, laboratory tests are used extensively in patient management in the following areas:

1. ***Diagnostic assistance.*** Laboratory tests are part of clinical findings that helps the clinicians in forming, and refining hypotheses. In some cases, a single laboratory result can confirm diagnosis. For example, diabetes is diagnosed based on an abnormally high blood glucose level. In other cases, laboratory results suggest certain diseases. For example, high blood urea content suggests renal failure.

2. ***Treatment planning and monitoring.*** Since disease severity may be determined directly by some laboratory tests, those results may be used to select therapy and to monitor progress. In the case of diabetes, blood glucose level determines the type of therapy used, ranging from diet management, to various oral hypoglycemic agents, to insulin treatments. Also, blood glucose level indicates the efficacy of the treatment regimen.

3. ***Routine health check.*** Many screening tests are performed for a selected section of the population to prevent serious illnesses. An example is the tumor marker PSA for prostate cancer.

4. ***Pre-operative anesthetic assessment.*** Prior to surgery, it is necessary to screen the patient for any hidden abnormality or borderline problems to prepare the medical team for any potential emergencies, since the body will be put under much stress both during and after surgery. This can be considered as a specific case of a routine health check.

4.2 The Mobile Scenario

The mobile version of the medical laboratory system is typically used when the physician is seeing the patient, either on rounds, or in the examining rooms. In this case, the patient and the chart are readily available, and the main goal of the task is to obtain the latest and/or the history of laboratory results for that particular patient. As a result, the emphasis of the system should be fast retrieval of information, and access of patient related information should only be available at the request of the user. This is in contrast to the desktop scenario where the system will be used away from the patient, at nursing stations, physicians' offices or even at physicians' homes. In that scenario, the patient or even the patient's chart may not be readily available when the physician views the laboratory results. Background information and context should therefore be provided on the screen as reminders.

4.2.1 Result Retrieval

It is likely that the patient's hospital number or health card number is available for fast and potentially more accurate patient identification in the mobile scenario. In some hospitals, patient's information is encoded in a barcode and thermally burned on a plastic wristband for positive patient identification. Since the physician only wishes to retrieve the result for one particular patient, input search can be based on a unique patient identification number to avoid ambiguity of spelling and difficulties in text inputs.

This contrasts with the desktop scenario where the patient's chart may not be available. In that case, it is unlikely that physicians would search using the patient hospital number. Also, it may be difficult for physicians to remember the exact spelling of the surname of all of their patients. Selection from a list instead of free input should be the used to initiate patient search.

4.2.2 Result Interpretation

Since the patient and the patient's chart are available, the default display should consist only of the results themselves. Background information, including patient demographics, contact information, and physician's notes, should only be available on demand to avoid unnecessary screen cluttering, data transmission and processing.

In the desktop scenario, since the patient may not be available to the clinicians at the time of result reviewing, it is particularly important to provide context for the user to facilitate result interpretations.

4.3 Conclusions from Tasks

Due to the particular demands of the mobile scenario, result retrieval can be initiated using a unique patient identification number to allow for fast and unambiguous retrieval. For interpretation, the emphasis is on the test values while patient information is available only on demand.

5 Presenting Laboratory Information

This chapter describes different aspects of presenting medical laboratory information. A study of laboratory data provides insights to the bounds of the problem at hand. The current paper-based report was used as a benchmark for the new design. Once the design goals are established, studies in human perception are used to derive design guidelines to achieve these goals.

5.1 Laboratory Data

In order to design a display system for numerical laboratory data, it is essential to understand the quantity and type of data to be displayed in laboratory reports. This profile was constructed based on existing paper-based reports.

5.1.1 Quantity

The total number of laboratory tests available to physicians differs from laboratory to laboratory. To obtain a sense of the potential quantity of information to be displayed, national standards are very useful. The American Medical Association annually publishes a listing of descriptive terms and identifying codes for reporting medical services and procedures performed by physicians called the Physicians' Current Procedural Terminology (CPT), which is used for billing in the United States. The Laboratory and Pathology section in the current CPT lists 287 tests. However, since many of them are test panels (i.e., groups of related tests), the true number of tests may be in the neighbourhood of 800 to 1000 (Warde Medical Laboratory, 2002). In practice, however, only about 50 tests are in common use (University of Wisconsin Family Medicine, 2002), even though the exact test selection differs from institution to institution. The number of tests in each laboratory report varies from 1 to about 30.

5.1.2 Type of Tests

Laboratory tests can be displayed using a number of organizational schemes. The simplest will be alphabetical. However, it is generally believed that grouping similar items in a hierarchical organization scheme facilitates visual search and reduces task complexity (Mullet and Sano, 1995). At the moment, laboratories organize the tests they provide by the processing procedures, similar to the classification used in the Laboratory and Pathology section of CPT-4[+] (Figure 1). Generally, there are three large categories of tests: Hematology, Immunology and Chemistry. However, from the physicians' point of view, the laboratory tests are performed based on the clinical question at hand, which may be better classified by anatomical or physiological systems, as shown in a popular medical reference (Figure 2).

Pathology and Laboratory
Organ or Disease Oriented Panels
Drug Testing
Therapeutic Drug Assays
Evocative/Suppression Testing
Consultations (Clinical Pathology)
Urinalysis
Chemistry
Hematology and Coagulation
Immunology
Transfusion Medicine
Microbiology
Anatomic Pathology
Cytopathology
Cytogenetic Studies
Surgical Pathology
Transcutaneous Procedures
Other Procedures

Figure 1. Listing from CPT-4: The Laboratory and Pathology section.

As seen in the listing from table of contents of CPT-4, laboratory tests are organized by the processing technique and science, rather than by body systems, which are more familiar and meaningful to most clinicians.

[+] CPT-4 (Current Procedural Terminology, 2004 edition)

1. Nutritional Disorders	12. Immunology; Allergic Disorders
2. Endocrine/Metabolic Disorders	13. Infectious Diseases
3. Gastrointestinal Disorders	14. Neurologic Disorders
4. Hepatic/Biliary Disorders	15. Psychiatric Disorders
5. Musculoskeletal/Connective Tissue Disorders	16. Cardiovascular Disorders
	17. Genitourinary Disorders
6. Pulmonary Disorders	18. Gynecology/Obstetrics
7. Ophthalomologic Disorders	19. Pediatrics
8. Ear/Nose/Throat Disorders	20. Disorders Due to Physical Agents
9. Dental/Oral Disorders	21. Special Subjects
10. Dermatologic Disorders	22. Clinical Pharmacology
11. Hematology/Oncology	23. Poisoning

Figure 2. A typical disease classification in medical references.

This classification of disease is almost universal among reference material for physicians. Note that it is very different from those used in laboratory medicine. This list is based on Merck Manual of Diagnosis and Theory

To study the type of tests ordered in typical clinical situations, two clinical cases are presented here as examples. The first case is a diagnostic workup of an infant presented with jaundice, while the second is a physical checkup of an elderly person otherwise healthy. The laboratory tests ordered by the physicians in charge are listed alphabetically in these case studies.

Case 1: Alagille's Syndrome (arteriohepatic dysphasia)

A six month old infant has been jaundiced since he was a neonate. On physical examination, he is a small, frankly icteric male in no distress. His height and weight are at the 5th percentile for age. His HEENT exam is benign other than his scleral icterus. His lungs are clear. On cardiac exam, he has a II-III/VI systolic murmur heard best in the periphery and on his back. He has normal heart sounds and good pulses. On abdominal exam, his liver is readily palpable 4 cm below the costal margin and it is quite firm. His spleen is readily palpable 3 cm below the left costal margin. The remainder of his examination is benign other than mild developmental delay and clubbing of his fingernails.

Work-up has included an abdominal ultrasound which demonstrated an echogenic liver without any evidence of intra or extrahepatic bile duct dilatation.

> Work-up has included an abdominal ultrasound which demonstrated an echogenic liver without any evidence of intra or extrahepatic bile duct dilatation. The gall bladder was of normal size and configuration. PRIDA scan demonstrated prompt excretion of isotope into the duodenum.
>
> His total bilirubin is 7 mg/dl with a direct fraction of 4 mg/dl. His ALT is 355 IU/L and his AST 297 IU/L. Alkaline phosphatase is 798 IU/L. He has a normal prothrombin time. His hematocrit is 32% and his platelet count is 110,000. Serum alpha-1-antitrypsin level is normal and hepatitis A, B, and C serologies are negative.

Note that the laboratory tests ordered are all related to the clinical case, namely tests directly related to liver functions (bilirubin, ALT, AST, alkaline phosphatase), indirectly related (blood clotting tests including prothrombin time, hematocrit, platelet counts), and popular alternatives like infectious diseases (hepatitis), and genetic disorders (Alpha-1 antitrypsin deficiency). One possible classification based on the Merck manual's table of contents may be hepatic/biliary, hematology, infectious diseases, and genetic disorders.

Case 2. Geriatric Health Assessment

> Jim Stokes, a 67 year-old white male, has come to the newly-established Seniors Health Center for his initial geriatric health assessment. Mr. Stokes was a smoker but currently does not use tobacco, rarely drinks alcohol, and has 3-4 cups of coffee each day. He does not exercise regularly. His diet is low-fat and high salt.
>
> Mr. Stokes reports a 20-year history of type 2 diabetes mellitus which is complicated by retinopathy, peripheral and autonomic neuropathy, coronary artery disease, and peripheral vascular disease. He's had progressive amputations of his toes (1985-6) and right forefoot (1988) at the V.A. Hospital. In addition, he had a four vessel coronary artery bypass operation in May 1995. Then in October 1995, because of chronic renal failure, Mr. Stokes reluctantly agreed to a kidney transplant. The donated kidney came from his son who was most insistent that his father accept it. In February 1996, he had a cholecystectomy.
>
> He had been hypertensive prior to his kidney transplant. Now he has orthostatic hypotension which is controlled by medication and a high salt diet. Also, because of the kidney transplant, he takes immunosuppresive drugs. Mr. Stokes indicated that he is allergic to penicillin and that Zest soap causes itching.

The following tests were ordered for his assessment:

Tests:	Patient's values	Normal values
Cholesterol	295 mg/dl	(140-200)
Triglycerides	427 mg/dl	(30-150)
HDL	28 mg/dl	(50-80)
Cholesterol/HDL	10.54	(5.00)
LDL	182 mg/dl	(< 160)
LDL/HDL	6.50	(3.55)
SMAC: Glucose	160 mg/dl	(70-125)
BUN	42 mg/dl	(10-26)
Creatinine	1.7 mg/dl	(0.4-1.5)
Heme Profile:		
RBC	4.0 x $10^6/mm^3$	(4.4-5.8)
Hemoglobin	13.1 g/dl	(13.5-18.0)
Hematocrit	37.2%	(41-53)

Table 2. Tests ordered for a Geriatric clinical case.

The tests ordered in this health assessment are mostly for providing baseline levels, and to screen for popular medical problems based on the patient's demographic group. In this case, the tests may be grouped under lipid profile (cholesterol, triglycerides, HDL, cholesterol/HDL, LDL, LDL/HDL), endocrinology (glucose), hematology (RBC, hemoglobin, hematocrit), and nephrology (BUN, creatinine).

5.1.3 Conclusions from Laboratory Data

In short, there is a large difference between the number of potentially available tests (in the order of hundreds), and the number of tests that actually appear on the report (typically below 30). As for the type of tests, most of them can be grouped under a handful of categories (3-5), either based on the processing technique and science, or based on anatomical or physiological systems. Due to the different reasons for ordering laboratory tests, and the degree of expertise of the clinicians in the clinical problem at hand, it is very difficult to predict the number of categories of tests that will appear on a "typical" laboratory report.

5.2 Existing Paper-Based Systems

Paper-based reports are the most common kind of laboratory results reporting system. Generally, each laboratory has its specific formats. Figures 3 and 4 are two examples from a local laboratory.

[redacted] *Page 3 of 3 #1178 E*

PATIENT [redacted] [redacted]

DATE OF COLLECTION **2001/02/19** PHN [redacted]

AGE **42 Years** SEX **M** DOB [redacted]

PATIENT'S PHONE [redacted] CHART NO

[redacted]

CC:
[redacted]

03C84

TEST	FLAG	RESULT	REFERENCE RANGE AND COMMENTS	
LIPIDS				
Cholesterol	H	**7.9**	(2.0 - 5.2) At risk:>6.2	mmol/L mmol/L
LDL Cholesterol	H	**6.2**	(1.5 - 3.4) At risk:>3.4	mmol/L mmol/L
		An LDL cholesterol level of less than 2.6 mmol/L is suggested for patients with established CAD.		
HDL Cholesterol		1.2	(> 0.9)	mmol/L
Chol:HDL Ratio	H	**6.6**	(< 5.0)	
		A ratio of 5.0 in males is associated with average risk. Increased ratios are associated with higher risk for CAD.		
Triglycerides		1.1	(< 2.3)	mmol/L
			Spec. Collected: 13.5 h p.c.	
THYROID FUNCTION				
TSH		2.07	(0.30 - 5.50)	mU/L
		If other thyroid tests are required after initial testing please phone 507-5070. Arun K. Garg M.D.		
TUMOUR MARKERS				
PSA		0.6	(< 4.1)	ug/L
REPORT COMPLETED				

Tests Requested:
FBS, HP, PSE, TRG, TSH, UR, UM, CHO, LDL, HDL

Printed: 2001/02/28 12:48 PATIENT [redacted]
Ordering Dr [redacted]
FINAL REPORT Page 2 DATE OF COLLECTION 2001/02/19

DR [redacted] CONSULTANTS IN LABORATORY MEDICINE

Figure 3. An example of paper-based laboratory results report showing the lipid profile and a tumor marker result.

[redacted] PAGE 2

Page 2 of 3 #1178 8

PATIENT [redacted] [redacted]

DATE OF COLLECTION **2001/02/19** PHN [redacted]

AGE **42 Years** SEX **M** DOB [redacted]

PATIENT'S PHONE [redacted] CHART NO

[redacted]

CC:

[redacted]

03C84

TEST	FLAG	RESULT	REFERENCE RANGE AND COMMENTS	
ROUTINE HEMATOLOGY				
HEMATOLOGY PANEL				
WBC		6.1	(4.0 - 11.0)	10*9/L
RBC		4.93	(4.30 - 5.90)	10*12/L
HGB		156	(135 - 180)	g/L
HCT		0.44	(0.41 - 0.52)	L/L
MCV		90	(80 - 100)	fL
PLT		167	(150 - 400)	10*9/L
DIFFERENTIAL				
Neutrophils		3.8	(2.0 - 8.0)	10*9/L
Lymphocytes		1.6	(1.0 - 4.0)	10*9/L
Monocytes		0.5	(< 0.9)	10*9/L
Eosinophils		0.2	(< 0.8)	10*9/L
Basophils		<0.1	(< 0.3)	10*9/L
URINALYSIS				
ROUTINE				
Appearance		Clear		
Colour		Yellow		
pH		6.5	(4.5-8.0)	
Protein		Negative		g/L
Glucose		Negative		mmol/L
Ketones		Negative		
Blood		Negative		
Leucocytes		Negative		
Nitrite		Negative		
GENERAL CHEMISTRY				
FBS		5.8	(3.6 - 6.1)	mmol/L

FBS Taken at: 1141 h 13.5 hrs.pc

Printed: 2001/02/28 12:48 PATIENT [redacted]

Ordering Dr [redacted]

FINAL REPORT Page 1 DATE OF COLLECTION 2001/02/19

DR [redacted] CONSULTANTS IN LABORATORY MEDICINE

Figure 4. An example of a paper-based laboratory result report showing a routine haematology panel, urine analysis and a chemistry result.

From the survey (Appendix A) conducted for this project and from results of a literature search, the advantages of the current paper report form include:

1. ***Brief and easy to read.*** It is difficult to find and interpret data if important information is clouded by irrelevant information, confusing groupings, and distracting fonts and colours (Politser, 1986; Wyatt & Wright, 1988).

2. ***Provision of standard values.*** Due to the many laboratory tests available, and the different units used in reporting results in different countries, it is essential that gender and age-specific standard values be provided for reference.

3. ***Familiar layout.*** Typically, physicians can glance at a report and acquired a fairly accurate picture of the patient's condition if they are familiar with the layout (Nygren et al., 1992).

Some disadvantages include:

1. ***Lack of Overviews.*** Most doctors found it difficult to gain an overview of the patient's clinical problems, past results, and treatment from the medical record (Wyatt, 1994).

2. ***Difficult to Compare.*** Comparison of data is hard. Data are spread across pages or record sections.

3. ***Difficult to Identify Trends.*** There are few graphs. Translation of dates into intervals is difficult.

4. ***Lack of Contextual Information.*** Generally, little patient-specific information is included in the paper form report (Wyatt & Wright, 1998). Individual laboratory results are difficult, if not impossible to interpret if they cannot be put in context, which includes patient-specific data such as sex, age, general physical characteristics (height, weight), blood pressure, temperature, previous medial history, and physician medical hypothesis (Goldschmidt & Lent, 1995).

5. ***Delivery delays.*** Generally, if the laboratory result is delivered via city postage, it takes three to seven days to reach the physicians. Tests performed in the hospital may have a

faster delivery rate, but there is still a delay. Emergency tests are usually delivered by telephone calls. The delay in delivery to the nursing station is therefore minimal during the operating hours of the laboratory (see the subsection *Lack of 365/24/7 availability* below). However, the physician in charge may not be aware of the delivery, and there may be transcription errors.

6. ***Lack of 365/24/7 availability***. When the laboratory is closed, test results will not be available or delivered, despite automated analyzers in some laboratories can be run overnight.

5.2.1 Conclusions from Existing Paper Systems

A network-connected result reporting system can overcome the problems of delivery delays and availability. However, perceptual problems of paper-based reports cannot be addressed by technology alone. In addition, by using a smaller display area, current perceptual problems may be exacerbated, while other problems may arise. Also, it is highly desirable to preserve the merits of a paper-based report. The next section looks at the laboratory report from the human perception perspective.

5.3 Human Perception and Laboratory Reports

Understanding human perception can help to identify design goals. According to (Politser, 1986), there are six natural stages of perception when viewing laboratory reports: filtering, simplification, coding, grouping, recognition and segregation.

1. ***Filtering***. Physicians screen data in a written laboratory report to locate key findings. Superfluous data are highly distracting, and filtering them may require extra cognitive efforts. Examples of superfluous data include unnecessary numbers and grids, repetitious reports of normal ranges, and unnecessary information.

2. ***Simplification***. Some laboratory data are reported to four digits, which is unnecessary. Also, data should be displayed in the simplest graphical format possible. In general, point graphs and bar charts are preferred over pie charts or Chernoff faces (Chernoff, 1973).

3. ***Coding***. To further reduce quantitative data, physicians tend to remember them qualitatively with respect to a reference point. For example, as above or below normal or being more or less than expected. Hence, providing the appropriate reference ranges in a laboratory report is extremely important.

4. ***Grouping***. Physicians tend to analyze data by grouping. This can be very difficult if the results are improperly arranged, when related results are shown on separate sheets of paper, or randomly grouped on the same sheet of paper. Examples include renal or liver function tests. In the case of renal function tests, if the temporal test results are not grouped together spatially, the synchronous patterns for serum urea nitrogen and serum creatinine are almost indiscernible. Missing this pattern may result in a faulty hypothesis of pre-renal disease, rather than renal disease.

5. ***Recognition***. At a later stage of perception, clinicians often must recognize less-obvious findings amidst irrelevant ones. This is a difficult task when unaided. For example, physicians may only need to monitor the change, rather than the absolute values of laboratory results. However, when the trend is subtle, it can be difficult to detect.

6. ***Segregation***. Physicians sometimes need to detect co-existing diseases. Intuitively, humans tend to separate the effects of co-existing diseases. However, in complex conditions involving multi-organ systems, the effect of one disease may conceal those of the others. For example, abnormal LDH isoenzyme profiles can be due to heart and liver problems. The concentration activities of five isoenzymes are usually plotted in a graph. Individually, heart and liver abnormalities results in different patterns on the graph. In the combined profile, the liver contributions mask those of the heart, since the resultant pattern is very similar to those based on liver contributions alone, despite elevated activity levels.

In general, an effective report layout should,

1. ***Organize data***. Human memory is best when the material is highly organized in a small number of clusters and in a hierarchical manner. Also, due to the limits in short-term memory, people cannot effectively handle more than ten pieces of information (or 7 ± 2

items (Miller, 1956)) simultaneously, and therefore optimization methods of data management are required. At the moment, most laboratory tests are classified by processing procedures in the laboratory. Since this classification scheme is also used in ordering the tests, this classification should perhaps be preserved in displaying the results. Nonetheless, confusions have arisen since physicians are generally not familiar with the basic procedures used in laboratory medicine. For example, from the questionnaire conducted for this thesis, one physician was confused about the organizational scheme used by the hospital since to her, the test IgA EBV could be under microbiology or immunology. A potentially better classification scheme may be the one employed by most medical references. However, consistency with existing ordering systems is more important in this case, since most of the time, the physicians who view the tests are those who ordered them in the first place.

2. ***Preserve patterns***. Consistent data organization on a page is important since familiar layout allows clinicians to access overall patient status without conscious effort (Nygren et al., 1992), and to make faster and more accurate decisions. With practice, physicians can encode the spatial positions of items in the report and map them onto meaningful data types, for example patient information and test types. In contrast, if the data is tightly packed, this mapping can be disrupted and results in a "pattern-dead" display. A special location on the report can be used excludively for patient demographics in order to provide a consistent pattern. As for the tests results themselves, due to the large discrepancy between the number of available tests and the number of tests on each report, it is impractical if not impossible to reserve spaces for each of the tests. However, it may be possible to create a fixed pattern of test categories in which the tests are grouped.

3. ***Highlight data***. Clinicians can quickly access the status of the patient if abnormal data is highlighted and "pops-out" with the first glance. A number of effective "pop-out" techniques have been reported, including colour, orientation, curvature, shape and so on (Ware, 2000, p. 151-200). Since the "pop-out" phenomenon is believed to occur in early vision, the data will be noticed at first glance, and do not need to be placed where they can be seen first. The use of highlighting can be very useful to give physicians an overall impression of the patient's status by highlighting abnormal results.

4. ***Provide Data on the Same Page***. Studies have shown that decision-making is faster and less error-prone if all data needed to support a decision can be viewed on one page instead of several sheets (Staggers & Mills, 1994).

5.4 Design Guidelines based on Perceptual Mechanisms

Due to the small size of many mobile devices, information display is a critical usability issue. A number of design guidelines were extracted based on researches on perceptual mechanisms to achieve the design goals identified in previous sections. In this section, six relevant mechanisms are discussed: rapid vision, perceptual organization, navigation, motion, perceptual layer, and attention and scene perception.

5.4.1 Rapid Vision: Highlight Data for Rapid Patient Status Assessment

Physicians screen data in a written laboratory report to assess patient status. Superfluous data are highly distracting, and filtering them may require extra cognitive efforts. If important data are "highlighted" for rapid vision processing, physicians can quickly access the status of the patient at first glance and with minimal effort. It has been suggested that rapid vision can assist in tasks for high-speed target detection, boundary identification, region detection and numerical estimation. Healey et al. (1996) shows that preattentive processing can be used for numerical estimation using colour and orientation independently. Similar visualization technique may be applicable to rapid patient status assessment.

Rapid vision occurs in the early feature abstraction stage where a visual image is analyzed based on primitive elements like form, motion, colour, and stereo-depth. Such preattentive processing performed in parallel and is rapid (less than 200 milliseconds). This early stage of processing is automatic, without attention, and thus does not require conscious effort. If information is encoded using a unique visual primitive that deviates from the rest of the elements on the display, the highlighted information will "pop-out", and reaction time to locate the highlighted object will be independent of the number of distracters present on the same display.

Not all properties of an object can be used to create this "pop-out" effect, and not all of these properties are perceived independently of each other. For example, it is relatively difficult to locate an L-shaped figure in a sea of "T"s, even though the target is unique in shape. Visual primitives that have been found to be preattentively distinct are colour, intensity, orientation, size (as in length), curvature, motion, and stereoscopic depth (Ware, 2000, p. 165). In order to be distinguishable at this early vision stage, the coding property in these visual primitives has to differ significantly. For example, to be preattentively distinct, a colour should lie outside the boundary of the region defined by all the other others in the local part of the display (Bauer et al., 1996). In other words, the highlighting colour should be outside of the convex hull of the colours in its neighbourhood as defined by the CIE tristimulus space. For orientation, a 30-45^0 separation between target and distracters is required. However, smaller differences in visual elements between target and distracters have been reported to be distinctive (Healey et al., 1996). Generally, it is easier to detect the presence of a primitive in a target, rather than the absence (Triesman & Gormican, 1988). Also, the distinctiveness of these primitives depends on the amount of feature variations of the distracters and the background.

Some features may interact with others to create an overall impression, instead of having individual preattentive effects. For example, width and height of the objects are integrated and interpreted together (integral features), while colour and texture are interpreted separately (separable features).

Before looking into how rapid vision can aid rapid patient assessment, a detour is made to review several techniques that have been used to display quantitative laboratory results. Generally, the display techniques can be classified as tabular or graphical.

5.4.1.1 Tabular Displays

Paper medical result reports use a tabular display format. Generally, it is the most space-efficient display format when compared to graphical plots, flow charts, or logical trees.

A number of issues have been raised regarding numerical laboratory result displays.

1. ***Non-standardized Use of Units***. Different countries and regions, and sometime even different laboratories within the same region, use different units to report numerical test results. Laboratories attempt to alleviate this problem by providing standard or normal reference ranges for each test. A simpler and perhaps more effective

reporting method is to report results in terms of deviation from the mean, as standard deviations (Mayer et al., 2001).

2. ***Significance of Abnormal Values.*** Studies showed that even in prestigious medical institutions, available laboratory data fail to elicit an appropriate clinical response in 30% of cases (Altshuler, 1994). Further, the same medical information may lead to different conclusions by different physicians depending on the knowledge and experience from the physicians, and the form and context of presentation (McNeil et al., 1982). Usually, clinical decisions are based on the distinction between normal and abnormal findings. The ability of physicians to interpret the varying degrees by which a result departs from the normal is therefore important. With the large number of tests and analysis techniques, the laboratory scientists and the pathologists are better suited to interpret laboratory tests than clinicians. One approach is to provide differential diagnosis (Werner, 1995), or some indication of the degrees of abnormality along with the numerical test values.

3. ***General Layout Issues.*** Familiarity or expectation has been cited as the most important consideration in displaying laboratory results. Nygren et al. (1992) argued that the structure and physical layout of medical records provide physicians with overviews of patient status at first glances. In fact, errors can arise when the display violates the readers' expectation about how the data are organized (Wright & Threlfall, 1980).

5.4.1.2 Graphical Plots

Graphical plots have been argued to be more effective in conveying trends than tabular displays, especially when decisions have to be made under severe time pressure (Schwartz & Howell, 1985). Two main display formats have been applied to laboratory results: single and multi-variate.

5.4.1.2.1 Single variate displays

Single variate displays show a graphical chronological history of a particular test result on a traditional X-Y plot, with X, the horizontal axis being the time line, and Y, the vertical axis being result values. There have been suggestions regarding the unit and scale of displaying laboratory results. Mayer et al. (2001) suggested a logarithmic time scale that provides detailed inspections of recent results, but retaining the older results in the "background". For the results, Mayer et al. (2001) suggested reporting standard deviations from the normal values, instead of the conventional numerical values so as to produce standardized and uniform plots. Powsner and Tufte (1994) proposed a similar technique to provide a standardized way for summing up patient status. In fact, standardizing graphical display can avoid errors related to differences in aspect ratio (Carmill & Thornton, 1992), and scale divisions (Chapanis, 1965).

5.4.1.2.2 Multi-variate displays

Multi-variate displays show a cross-sectional snapshot of a number of test results at any given time in a patient's medical history. In current paper reports, these snapshots are displayed as tables (Figure 5), sometimes using an abnormal value indicator (Figure 7). The most common abnormal value indicators are colour and special characters (for example, asterisks). Graphical displays generally take the form of polygon-based displays (Figure 6) or bar chart-based displays (Figure 8).

Polygon-based. One example is shown in Figure 6. Each test's axis radiates from a common center, with the test name and numerical result printed adjacent to the outer end of the axis. Result mapping is done according to a reference table, so that all the test results can be displayed on the polygon. Different regions of the polygon are colour-coded to indicate the various degrees of deviation from normality, including "Dangerously Low", "Low", "Uncertain Low", "Normal", "Uncertain high", "Elevated" and "Dangerously elevated". (Hooke et al., 1991).

Bar chart-based. This has the general appearance of a traditional bar chart (Figure 8). Test names (e.g., 1-L, 2-M) and results (e.g., 750, 53.7) are printed below the plot. Test results are plotted using black horizontal lines, and are connected by thin, vertical black lines. Similar to the polygon-based presentation technique, different regions of the bar chart are colour-coded to indicate the various degrees of deviation from normality.

	06/01 09u00
1-L	750
2-M	53.7
3-S	5.44
4-L	721
5-M	48.8
6-S	5.15
7-L	748
8-M	75.9
9-S	1.90
10-L	439
11-M	69.2
12-S	3.01

Figure 5. A monochrome, tabular representation.

The labels (e.g., 1-L) are listed on the left hand column while their corresponding numerical values (e.g., 750) are listed on the right column of the table. Separators are inserted for readability and do not infer grouping.

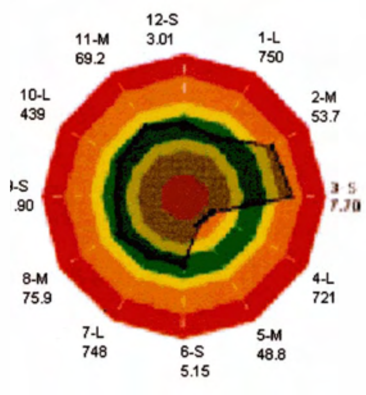


Figure 6. Polygon-based display.

This is a multi-variant graphical display technique in which each test's axis radiates from a common center, with the test name and numerical result printed adjacent to the outer end of the axis, with the black dots indicating the actual data points on the plot, and black lines linking the data points. Result mapping is done according to a reference table, so that all the test results can be displayed on the polygon. Different regions of the polygon are colour-coded to indicate the deviation from normality of each region, including "Dangerously Low", "Low", "Uncertain Low", "Normal", "Uncertain High", "Elevated" and "Dangerously elevated". The area enclosed by the data points is shaded for better readability. Based on Hooke et al. (1991).

	06/01 09u00
1-L	750
2-M	53.7
3-S	5.44
4-L	721
5-M	48.8
6-S	5.15
7-L	748
8-M	75.9
9-S	1.90
10-L	439
11-M	69.2
12-S	3.01

Figure 7. A coloured tabular representation.

Cells in the table are colour-coded to represent the urgency of the result.

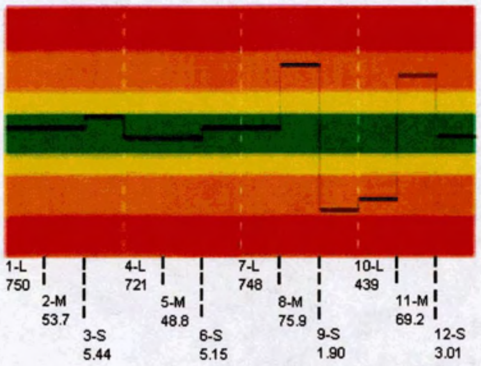


Figure 8. Bar chart-based display.

This has the general appearance of a traditional bar chart. Test names and results are printed at the bottom of the plot. Test results are plotted using black horizontal lines, and are connected by thin, vertical, black lines. Similar to the polygon-based presentation technique in Figure 6, different regions of the bar chart are colour-coded to indicate the various degrees of deviation from normality. Based on Hooke et al. (1991)

Verheij et al. (1997a, 1997b) showed that tabular displays with colour-coded cell background (Figure 7) are found to be more efficient and accurate in assessing overall patient status than multivariate graphical displays (Figure 6, Figure 8).

The results obtained by Verheij et al. (1997a, 1997b) may be partly explained in terms of rapid vision. In order to determine overall patient status using the monochrome tabular display (Figure 5), users have to read individual test entries and compare the result to the corresponding nominal ranges. In both the colour-mapped tabular and graphical displays, this laborious and error-prone classification process is replaced by various urgency class indicators. In the bar chart-based (Figure 8) and polygon-based (Figure 7) displays, visual search is not effectively guided for rapid status assessment for a number of reasons. First, colour is used in the scale to indicate different urgency classes, rather than to highlight abnormal data. As a result, the overall effect is more distracting than informative. Second, since the black line links individual test results, it is more effective in encoding relative, rather than individual test status. This makes it more difficult to locate the abnormal test results. Third, the black lines in the graphical displays visually link test results that may or may not be related, thus creating perception of non-existing trends. In contrast, the tabular background-coloured presentation technique highlights the abnormal results. It is perhaps due to this reason that the tabular displays which highlighted abnormal values performed better than the graphical ones. In fact, most of the errors made in the graphical displays, and in the monochrome tabular display were missed abnormal results (Verheij et al., 1997b).

In conclusion, cross-sectional, snapshot quantitative test results are best displayed in tables, due to the familiarity and expectation of physicians in receiving this information, and to the questionable effectiveness of multi-variant graphical displays. Highlighting abnormal test results using colour can take advantage of the capability of vision, and provides rapid patient status assessment.

5.4.2 Perceptual Organization: Organize Data as Diagnostic Assistance

The Gestalt law of proximity states that elements that are closest together will be perceived as belonging together. This visual organization may help users in data analysis by helping them structure their mental organization, and thus facilitates visual search, decision-making and reduces errors, since memory is best when the material is highly organized in a small number of clusters and in a hierarchical manner.

One obvious area for proximity grouping in a laboratory report is patient-specific and test-specific information. Typically, patient-specific information is placed at the top left-hand corner of the report, while test-specific information occupies the rest of the space.

For test-specific information, further grouping is required since it is unlikely that all the test information will fit on a single screen, especially for small screens. Data analysis is particularly difficult when the results are improperly arranged, when related results are shown on separate sheets of paper, or randomly grouped on the same sheet of paper. Examples include renal or liver function tests. In the case of renal function tests, if the temporal test results are not grouped together spatially, the synchronous patterns for serum urea nitrogen and serum creatinine are almost indiscernible. Missing this pattern may result in a faulty hypothesis of pre-renal disease, rather than renal disease. In fact, studies have shown that decision-making is faster and less error-prone if all data needed to support a decision can be viewed on one page instead of several sheets (Staggers & Mills, 1994).

Despite the apparent benefit of grouping results, results from medical cognition cast doubt on the use of grouping under all conditions. Laboratory results are usually used as part of clinical findings available to physicians for decision-making. Medical problems are complex problems, and clinicians tend to be heuristic rather than algorithmic in their approaches (Patel et al., 1994). In addition, their reasoning approaches differ depending on their level of expertise, and their domain knowledge (Arocha et al., 1993; Patel et al., 1994; Patel et al., 1997). In order to facilitate effective problem solving for all users, the system needs to be flexible enough to accommodate these differences. For experts, laboratory tests are used to refine and evaluate their hypotheses, and to plan therapy. Since each test is ordered for a reason, experts only need to be shown the test results, which will be interpreted in the framework of the hypotheses. In this case, results should be displayed in a simple and straightforward manner to avoid grouping ambiguities. One suitable presentation technique would be an alphabetically sorted list. For subexperts and novices, since their reasoning may not be structured and coherent, laboratory tests may be used to search for clues and thus generate more hypotheses, rather than to evaluate existing hypotheses. In that case, it would be helpful to structure the tests to facilitate a more structured reasoning.

Even when the designer is certain about the benefit of grouping, he still needs to decide on the organization scheme. Issues regarding the current grouping scheme, and the proposed anatomical/physiological scheme were described in §5.1.2.

While hierarchical proximity grouping aids diagnosis by putting similar results close together spatially, the principle of similarity helps the physicians in linking between these groups. One obvious use of similarity is to highlight abnormal test results by colour. If a single colour is used, abnormal results will be perceptually grouped together despite the lack of spatial connection. This may be very useful for diagnosis of multiple organ diseases (e.g., renal problem leading to heart failure), as clinicians can quickly locate abnormal test results from multiple categories despite the spatial separation.

5.4.3 Navigation: Detail-on-Demand to Overcome Small Screen Size

The small screens limit the amount of information displayed on each screen. As a result, in order to display complex data on a single screen, designers either reduce the font-size or employ scroll-bars. Alternatively, data will have to be spread over a number of screens. Both of these techniques have been shown to decrease readability (Duchnicky & Kwahk, 1983), increase search efforts (Jones et al., 1999) and cause loss of orientation (Dillon et al., 1990).

Laboratory results are essentially a linear list displayed in a table. The length of the table is determined by the total number of tests in a report. The width of the table is determined by the number of fields, or attributes, associated with each test. A typical list includes the test name, the numerical result, the nominal reference range, the unit, the date of sampling and notes from the consultant and/or the pathologist.

It is immediately obvious that both the length and the width of the table are beyond the display capability of a typical small device. Visualization techniques are therefore important to avoid excessively scrolling, or spreading the table over too many screens. One of these techniques is "detail-in-context". Despite the benefit of showing interesting areas in detail without losing the background context, "detail-in-context" visualization techniques can be difficult for small screens. First, most of their implementations require substantial processing, which will render the application too slow to use on small devices. Second, these techniques tend to suffer from screen cluttering, even with desktop sized display. With the small display areas on mobile devises, the problem of screen cluttering may be more difficult to resolve.

Three feasible visualization techniques for small screen are elision, Table Lens distortion and a more restrictive form of graphical distortion.

5.4.3.1 Elision Techniques

Since users can only attend to one object at a time, unattended objects can be hidden until requested. This "detail-on-demand" technique is similar to the cognition process of chunking in the sense that unattended objects are grouped into a single object (Ware, 2000, pp. 252-255). Examples include the intelligent zoom system (Bartram et al., 1994), NV3D system (Parker et al., 1998) and CZWeb (Tan, 1997).

In the result reporting application the elision technique is useful in providing contextual and detailed information on demand. For a laboratory test report, context is defined as the actual physical state of the patient and the set of reasoning that the physician is considering in the evaluation of the patient at a specific point of time. Each test result should be placed within the context in order to be meaningfully interpreted. Along with the physician's medical hypotheses, the following information will help putting results in context,

1. ***Patient-specific data.*** Sex, age, general physical characteristics (e.g., height, weight).

2. ***Current medical status.*** Blood pressure, temperature.

3. ***Reasons for ordering test.*** Examples include monitoring, routine health check, and diagnosis.

All of this information, despite being important in putting the test results in context, is not required at all times and should therefore only be available when requested. For example, the physician's note provides important context and clue for interpreting laboratory tests. However, it is usually not required for initial analysis of the laboratory data. One way to work with the constraints of the small screen is to only show a "note indicator" (e.g., the "[]" icon in Figure 13) in the table to allow for access to its corresponding physician's note, which will be displayed at the bottom of the screen upon user request. This effectively reduces the required width of the table.

The second area where the elision technique can be used is to filter out uninteresting laboratory results. Since laboratory tests are used to rule in and out hypothesis, physicians tend to investigate a number of hypothesis at the same time. Once used to rule out pathology, the previously useful tests are distracting and take up valuable screen space. Also, there is a need to provide related results on the same screen. To facilitate that, uninterested tests should be hidden. One method to achieve this goal is by elision.

Before this technique can be applied, the linear list has to be converted to a hierarchical tree structure. As seen in §5.1.2, there are a number of issues in grouping scheme selection, but such conversion facilitates navigation by hierarchical grouping. According to Furnas (1999), hierarchical grouping and fisheye samplings have better navigability than simple linear structures, in terms of two quantitative measurements: efficient view traversability (EVT) and view navigability (VN). EVT is expressed in terms of the maximum number of out-going links and the diameter of the viewing graph, and both are superior in hierarchical structures. In terms of VN, labeled hierarchical trees offer better navigability than its linear list counterpart, especially when the organizational scheme is presented to the user as context.

Once converted, laboratory tests in a report can be grouped under a handful of categories (about 3-5), which can be displayed on a single screen without vertical scrolling. This implementation is similar to the graphical file-structure provided by most operating systems and the accordion summarization technique devised by Buyukkokten et al. (2001). Upon user request, the selected category heading will expand and display all the tests grouped under that specific category.

In addition to providing a means to enhance visual search, the hierarchical + elision presentation technique also provides a patient status overview. When visible, the category headings serves as a bird-eye view to the nature of the tests ordered, thus providing the physicians with the type of disorder in question. Also, if the abnormal headings are highlighted, the physician can further access the status of the patient at this stage, and may help to direct his attention to the problem areas.

An alternative implementation of the elision technique is to use fisheye sampling (Furnas, 1999). Instead of geometrically distorting the display, which is computationally expensive, fisheye sampling displays a selected portion of the tests based on the area of interest. The visibility of each test is a function of its distance from the selected test. In other words, from the currently viewed test (say test 1), only tests 2, 4, 8, 16… in the list will be shown. In other words, the local part of the list is shown in more detail, while keeping the farther away tests as cues for context. In this way, the total number of tests shown can be fitted on the screen. This technique can be used to display tests alphabetically.

5.4.3.2 Distortion Technique—Table Lens

In Table Lens (Rao & Card, 1994), the whole table is "squeezed" into the available space, with row widths, or column heights, or both, altered when necessary. Table Lens is an example of distortion technique where all the information in the visual space is displayed on the screen.

Local expansion of the display allows the user to see a portion of the visual space in more detail, while the rest of the display contracts accordingly. As a result, the user can view a selected area of interest without losing the surrounding context. Other examples of distortion include the dynamic fish-eye (Lamping et al., 1995), distorted map (Leung et al., 1994), and continuous zoom (Schaffer et al., 1996).

As a result of displaying the entire table on the screen, not all the cell contents in the table are completely visible. Upon user request, the cell of interest will expand to its full size, and its neighbouring cells will shrink in size to accommodate its expansion. Since this adjustment involves all of the cells in the table, it is computationally expensive for small devices. In addition, due to the large number of columns required in displaying laboratory test results, only three columns can be fully visible at best, thus severely reducing the ability to provide enough context for orientation.

Due to these limitations, the elision technique is selected for implementation of the laboratory result reporting system.

5.4.3.3 Distortion technique—Two views of a graph

While most of the implementations of the "detail-in-context" technique address spatial problems (for example, geographic maps, data and navigation maps, and calendar), structural and temporal scale problems can also be addressed using the same technique. An obvious structural example is computer software source code. Complex computer systems are typically organized in many layers, with module, classes and components. "Detail-in-context" visualization technique can be used to view a single function (detail), in the local context of its module and the global context of its layer. Temporal viewing of historical events is an example of applying "detail-in-context" technique to temporal problems.

Physicians often need to study temporal development of laboratory results, either to discover trends of disease progression, or to monitor treatment effectiveness. Graphical displays have been argued to be more effective in conveying trends than tabular displays, especially when decisions have to be made under severe time pressure (Schwartz, 1985).

For the time scale, Mayer et al. (2001) suggested a logarithmic time scale to provide detail inspections of recent results, but retaining the older results in the "background". Despite its benefits, it is more difficult to accurately discern trends in a logarithmic time scale. Ideally, the users should be able to take advantage of both time scales. This may be feasible by employing distortion techniques, if the user can navigate between the two views smoothly. In other words, the logarithmic time scale will be used to obtain a more detailed view of recent data, while the linear time scale view will be used for a more accurate assessment of trends. Since only two views are possible in this case, it is a considerably more restrictive application of the "detail-in-context" technique on a temporal problem, but precisely because of this restriction, this visualization implementation may be more feasible on a CPU-limited mobile device.

One disadvantage of using distortion technique is the constantly changing visual pattern. As mentioned before, the ability to preserve pattern is one of the merits of paper display. However, with the small screen, it is virtually impossible to preserve pattern in the test result area. However, consistent patient specific information and the fixed-patterned anatomical overview discussed below (§5.4.6) may help to alleviate this problem.

5.4.4 Motion: Object Constancy

As discussed in §5.4.3 "detail-in-context" visualization techniques may be helpful in displaying laboratory results on a small screen device (elision in §5.4.3.1; distortion in §5.4.3.2 and §5.4.3.3). For the elision technique, since a potentially large number of rows can be inserted into the table, animation will avoid disorientation by showing the intermediate steps of the transformation, thus providing continuity of the display and allowing constant update of the users' mental representation. For the two-view graphical display, animation could link the linear and logarithmic graphs. Rather than to provide orientation, the purpose of animation here is to relate the two graphs spatially, since most physicians are not familiar with logarithmic scales. In other words, animation will help users to interpret the unusual logarithmic time scale and to avoid possible misinterpretations due to the "distorted" scale.

Indeed, animation is an essential component in many "detail-in-context" visualization techniques since it is believed to support object constancy. According to Robertson et al. (1989):

> An important property of interactive animation is that it can shift a user's task from cognition to perceptual activity, freeing cognitive processing capacity for application tasks. For example, interactive animation supports object constancy. Consider an animation of a complex object that represents some complex relationships. When the user rotates this object, or moves around the object, animation of that motion makes it possible (even easy, since it is at the level of perception) for the user to retain the relationships of what is displayed. Without animation, the display would jump from one configuration to another, and the user would have to spend time (and cognitive effort) re-assimilating the new display. By providing object constancy, animation significantly reduces the cognitive load on the user.

In almost all current "detail-in-context" implementations, smooth animation that links the image between transitions is an essential requirement for effective visualization, and object constancy is assumed to be the reason behind the requirement. Fisher and Dill (2000) postulated that in a dynamic display, smooth animation is required for continuous processing of the objects in an image so as to "keep track" of them. This processing is believed to be low-level, involving a number of spatial pointers to the objects individuated for cognitive processing. When the features of these objects change (in this case, their locations), their mental representations, or object files, will be updated as long as the pointers are intact. In other words, smooth animation in dynamic display is required to preserve these pointers.

5.4.5 Perceptual Layers: Overcoming Small Screen Size

The Perceptual Layering methodology can be used to alleviate the constraints of small screen size and resolution on mobile devices, especially for tabular displays (Van Laar, 1999). Layering reduces visual clutter, as information is perceived as being on different layers. User's attention will be directed from the most important to the least important objects on the display due to the illusion of depth difference produced by the layering. As a result, the designer can put potentially more information on the small screen.

The use of perceptual layers has been advocated as a means to organize task relevant data elements on maps and displays (Wood, 1968; Van Laar 2001). For example, in a map, motorways and main roads can be organized into separate layers from land use or minor roads. The same principle is used in the Perceptual Layering methodology devised by Van Laar (1999; 2001) to organize information, direct attention and maximize the use of space on computer displays. Using principles of colour and depth perception, the display is structured into perceptually different layers.

The process begins with task analysis to determine the structure of the task in relation to the information to be displayed. Elements within the display are first analyzed, and ranked according to their importance in the task supported by the display. Elements that are ranked at the same level are grouped together in the same visual layer.

Robinson et al. (1995) describe three ways in which visual layers can be used to communicate this structure. First, stereogrammic layers direct attention to the most important informational layers through figure-ground effects. Second, extensional layers represent quantitative relationship between layers, and third, subdivisonal effects are used to communicate relationship within a layer.

In Van Laar (2001)'s methodology, colour is used in creating these perceptual layers for its diversity, since the designer can manipulate the three basic axes of colour: hue, saturation and lightness. However, the same principle can be applied to monochrome displays without using hue.

Using this visualization technique, information on the display is grouped in two dimensions: by layer (saturation and luminance), and by task structure (hue). Layering helps the interpretation process, as similarity among information is encoded as being on the same perceptual layer. Colour-coding helps to locate information, as users can perform hierarchical based on task structure, instead of linear search of the whole display. Also, if the hierarchical organization of the information matches the users' expectation and knowledge, users can apply their categorical and semantic knowledge to further speed up the search process.

Laboratory results are generally displayed in tables. Using the Perceptual Layering methodology, information can be grouped as follows,

Perceptual Layer	Information Groups	Justification
1 (lowest)	Background information for the tests, including the nominal range, the unit, and the date of sampling. Patient demographics also belong to this group.	These are less important information in the report, since they are mostly static from report to report for the same test, and are provided for reference
2	Categorical headings	Categorical headings provide visual guide in searching. They are not static from report to report.
3	Dynamic information for the tests, including the test names and the numerical results	Test results and names are the task-important information that users need to locate, compare and recall, and should be easily seen.
4 (highest)	Abnormal test results	Abnormal values are highly important information and should be most conspicuous.

Table 3. A possible layering scheme of the laboratory report for perceptual layering.

As for grouping by hue, the two obvious areas are patient specific- and test specific-information. This is similar to, and may enhance, the proximity grouping technique as discussion in §5.4.2 where patient-specific and test-specific information are spatially grouped for ease of interpretation and visual search. In addition, the different perceptual layers may enhance the perception of the hierarchical structure discussed in the same section.

5.4.6 Attention and scene perception: Object Display

Despite the limited capacity of attention, parts of an attended object (or proto-objects) can be accessed quickly (Rensink, 2002). In other words, if important data is grouped into a unit of attention (i.e., a coherence object), users can quickly access its property (i.e., its proto-objects) quickly. This idea is similar to that of an object display (Wickens, 1992), where a large number of separate variables are displayed using a "single contoured object". The advantage of such displays is three-fold. First, by putting these variables into a single object, its processing will be processed together in parallel. Second, the compact object display can reduce visual cluttering. Third, parts of this attended object can be accessed quickly. One example of an object display of multivariate data is the Chernoff faces (Chernoff, 1973), which use features of the human face to encode data. For example, the level of parameter 1 can be encoded using the degree of curvature of the mouth, and parameter 2 can be encoded using the height of the ear. Despite its ability to group data into a single object, Chernoff faces may not be appropriate in displaying multivariate data. First, our sensitivity to differences in facial features is highly nonlinear. For example, we may be more sensitive to curvature of the mouth, than the height of the ears. Second, facial features may interact and produce unexpected perceptive results. Third, reaction and interpretation of facial features is highly individual.

To truly harness the potential benefit of object displays, it is better if the object resembles the data it is representing. In the laboratory result reporting application, one of the object candidates is the human body (Figure 9), since in most cases, individual laboratory test indicates the functional status of a system in the body. Using the human body object display, each system is displayed as its corresponding anatomical counterpart in the object. Based on the results of the associated laboratory tests requested, the functional status of these systems can be indicated in a binary fashion: alarm (highlighted in red), or normal (not highlighted). Access to the individual tests and their numerical results can be revealed upon user request (Figure 10).

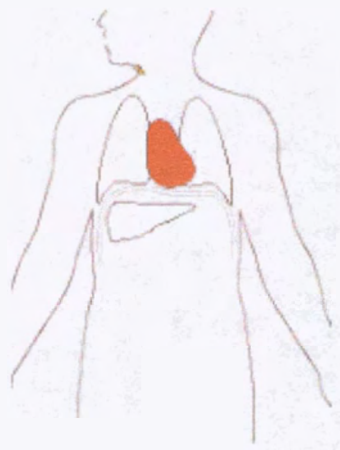


Figure 9. Using the human body as an object display for laboratory tests.

Laboratory test are grouped into functional or anatomical systems, and the status of these systems are shown as alarm (highlighted), or normal (not highlighted). In this example, the cardiovascular subsystem is highlighted, while others (respiratory system, liver, and thyroid) are investigated, but not found to be functionally abnormal.

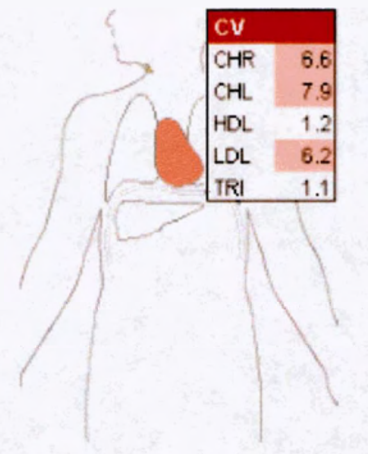


Figure 10. Viewing of individual tests in a whole-body display.

Users can request further information from any system. Upon request, tests that are related to the system will be displayed. In this example, tests belonging to the cardiovascular system are displayed.

5.5 Conclusions from Presenting Laboratory Data

Despite the large number of laboratory tests available for physicians, only about 50 tests are in common use, and these tests can be grouped into six or less categories. In order to present test results effectively on a small screen, perceptual mechanisms were investigated to derive design guidelines for the prototyping process. The mechanisms considered in this chapter include rapid vision, perceptual organization, navigation, motion, perceptual layers and scene perception.

6 Prototype Design Process

The initial prototype design was developed based on the design goals and constraints identified in the previous chapters: characteristic of the target users, the main supported tasks, the limitation of the target devices, and studies on human perception. The initial design was evaluated with eight physicians in an observational user study. Their comments and reactions helped to develop the second prototype, which was then evaluated by a sub-group of the initial study group. Further designs were incorporated to develop the third and final prototype.

6.1 Initial Design: Design Version 0

The main tasks of a medical laboratory results reporting system are patient status assessment, and diagnostic and treatment decision support. The following is a summary that lists the design issues and their corresponding presentation techniques in these two tasks.

6.1.1 Patient Status Assessment

6.1.1.1 Design Issues

Physicians screen data in a written laboratory report to assess patient status. They should be able to do so quickly, accurately, and without effort. In order to satisfy these requirements, abnormal test results should be perceptually different from other information on the display. For small screens, an additional design issue is the limited screen size. In many case, not all of the results can be accommodated on a single screen, making rapid assessment difficult.

A related problem is the need to preserve a consistent pattern on the display report, so that physicians can quickly assess the status without detailed study. This is particularly difficult on small displays since the expensive screen real estate is devoted to displaying test results, and the number and type of tests requested differ from report to report.

6.1.1.2 Presentation Techniques

Two presentations were used at different points of the application to solve these issues: the anatomical and the hierarchical overviews. These two displays are used under different situations. The anatomical overview is used purely for status assessment, while the hierarchical display is also used for more detailed diagnostic and treatment planning support.

6.1.1.2.1 Anatomical Overview

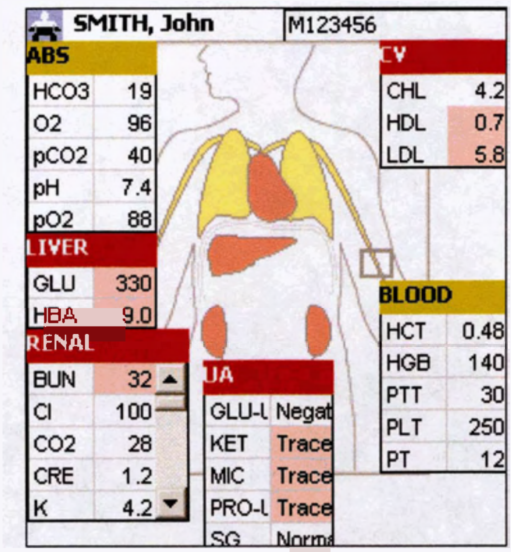


Figure 11. Anatomical Overview

The Anatomical Overview (Figure 11) is designed to allow quick assessment of the patient's status based on anatomical systems. Using this pictorial presentation has five advantages.

Familiar Organization. The anatomical system is familiar to all physicians. Using this organizational scheme reduces the learning required before using the system.

Constant Positioning. The constant anatomical map provides a pattern for physicians to quickly assess the patient status, or to ascertain the status of a certain organ system.

Compact Hierarchical Display. The pictorial display provides a summary of the laboratory tests by grouping the tests under a compact hierarchical display. In other words, despite not being able to show all the individual tests (some can only be revealed by scrolling in the tables), physicians can still obtain a sense of the patient status by the pictorial display.

Highlight Abnormal Results. Organ systems under investigation are shown with the abnormal systems highlighted for rapid assessment. Similarly, abnormal tests are also highlighted for to allow for the "pop-out" effect in rapid vision.

Fast Acquiring of Information. Since information is associated with a single object, it is believed to assist information pickup based on the object display model.

Once the area of interest is located, the user can select either the test, or the organ system to access detailed information displayed in the hierarchical screen, which will be discussed in detail in §6.1.2.

6.1.1.2.2 Hierarchical Organization of Tests

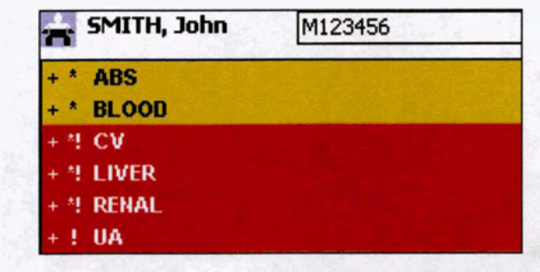


Figure 12. Hierarchical Overview

Individual laboratory tests are grouped under the category headings as shown in the figures. "+" indicates the category state as being "closed", when the individual tests grouped under the category are not visible. "*" indicates the tests in that category have not been viewed, while "!" indicates alarms or abnormal test result(s) within the category.

This presentation technique (Figure 12) is similar to that of the Anatomical Overview. However, instead of displaying the results pictorially, only the headings and their status (alarmed versus normal) are displayed.

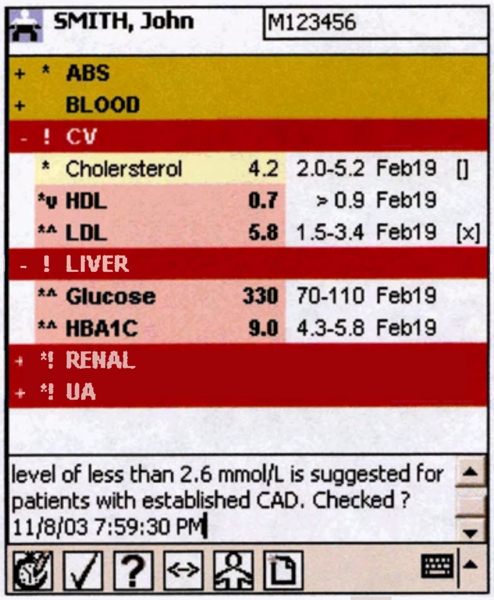
6.1.2 Detail Diagnostic and Treatment Planning Support

6.1.2.1 Design Issues

The second purpose of the system is to allow the physicians to perform a detailed study of the laboratory tests for diagnosis and treatment planning. There are at least three aspects of this activity. First, in order to be an effective tool, the system needs to support the decision-making process of physicians. It is known that physicians build hypotheses regarding the clinical case at hand and organize available information into structures relevant to those hypotheses. The organizational scheme of the test results should not impede this decision-making process. Second, since it can be difficult to display all test results on a single screen, it is essential to minimize the amount of "cross-screen" referencing required to making decisions. Third, physicians may require supporting information like patient demographics, nominal values of tests, specimen collection date, and any notes provided by either the laboratory, or referring physicians.

6.1.2.2 Presentation Techniques

Two presentation techniques were used in presenting detailed test results: hierarchical with elision (Figure 13) and a linear display (Figure 14).



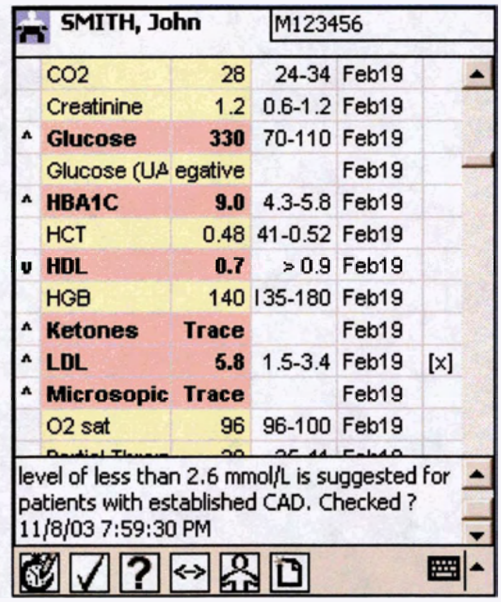


Figure 13. Hierarchical Display of Laboratory Tests.

Figure 14. Linear Display of Laboratory Tests.

"+" indicates the category state as being "closed", when the individual tests grouped under the category are not visible, while "-" indicates the state of "open", when tests are visible. "*" indicates the tests in that category have not been viewed, while "!" indicates alarms or abnormal test result(s) within the category.

For both displays, "^" indicates an elevated test value when compared to the reference range, and "v" indicates a low test value. Abnormal values are highlight by both colour and font (bolded) in the columns of test name and test value. The presence of notes from the laboratory are indicated by the symbol "([]" and annotated notes by the clinicians are indicated as "[x]".

6.1.2.2.1 Hierarchical + Elision Display of Tests

In this display, laboratory tests are organized by a system similar to the current paper-based system, except that tests that belonged to the large general group of "Chemistry" were extracted into related physiological or anatomical system. This decision was made based on the organizational scheme used by most popular medical references and the belief that such a familiar grouping scheme will better support the diagnostic process than the one currently used in laboratory medicine. Also, since the pathology of most diseases tend to localize at the organ level, it is likely that grouping of test results using this organization scheme will put pertinent tests physically close together, thus avoiding excessive "cross-screen" consultations during the decision making process. Figure 13 is a screen capture of the design, where laboratory tests were organized into systems like Arterial blood gas, Blood, Cardiovascular, Liver, Renal, and Urine Analysis.

Using a hierarchical display technique with elision in this application provides detail-in-context to overcome the difficulty in displaying the entire report on a single screen due to the small screen size. The hierarchical + elision technique provides a tool for the physicians to hide less interesting tests, so as to allow sufficient viewing area for the tests of interest without losing sight of the overall status of the hidden tests. Also, interesting test groups have a better chance of being viewed together on a single screen than otherwise when the hidden tests free up valuable screen space.

In addition to being a potential solution to the small screen size design challenge, the "detail-on-demand" technique may also be a workaround for the constraints of connectivity imposed on small devices. Since the entire report is not displayed initially, the report can be downloaded in multiple sessions, and only when requested by the users. This helps to spread out the data transmission "dead-time" and may provide more fluid interactions for the users.

Animation is believed to be an essential component in distortion techniques (Robertson et al., 1989). Due to the limited processing power of the small device, the process of inserting and removing rows from the data table is slow. In short, animation is provided automatically.

6.1.2.2.2 Elision Technique

Similarly, the use of the elision technique by itself also provides detail-on-demand. One example is the test "note". This information is important to physicians in interpreting the test results as it may be alert or extra information provided by the laboratory regarding the test, or clinical notes by the referring physician, or comments made by consulting physician who has reviewed the test.

In the current paper-based reports, this information is typically displayed below the test record, making the report difficult to read since the intervening note destroys the visual alignment and flow of the display (Figure 15).

LIPIDS		
Cholesterol	H	**7.9** (2.0 - 5.2) mmol/L
		At risk:>6.2 mmol/L
LDL Cholesterol	H	**6.2** (1.5 - 3.4) mmol/L
		At risk:>3.4 mmol/L
		An LDL cholesterol level of less than 2.6 mmol/L is suggested for patients with established CAD.
HDL Cholesterol		1.2 (> 0.9) mmol/L
Chol:HDL Ratio	H	**6.6** (< 5.0)
		A ratio of 5.0 in males is associated with average risk. Increased ratios are associated with higher risk for CAD.
Triglycerides		1.1 (< 2.3) mmol/L
		Spec. Collected: 13.5 h p.c.
THYROID FUNCTION		
TSH		2.07 (0.30 - 5.50) mU/L
		If other thyroid tests are required after initial testing please phone 507-5070. Arun K. Garg M.D.

Figure 15. A portion of the paper-based laboratory report shown in Figure 3.

It is therefore undesirable to display test notes in the main test area. Instead, the existence of notes is indicated by the symbol "[]", while the note itself is displayed at the bottom of the test area only by request (Figure 13, Figure 14). When it is no longer required, the note will be closed to increase the viewing area for tests. Similar to their paper counterpart, physicians can add their annotation beside the tests. In small devices, this is more difficult. To reduce the impedance to annotation, some frequently used phrases are pre-programmed and can be inserted by pressing the corresponding icon. Examples include for a time-stamp; for the word "checked"; and for the question mark "?". Once marked, the symbol for the note is changed to "[x]".

Another example of elision is the design choice to hide the unit of the test results. Due to the limited columns available for display, it is impossible to display both the specimen collection date and the test unit in the table without using horizontal scrolling. Given the test reference range, the unit provides very little clinical relevance in interpreting the test results, and is therefore hidden unless requested (Figure 16).

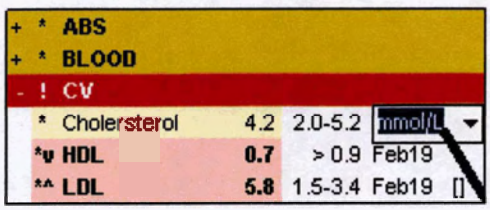


Figure 16. Screen shot showing that test unit is hidden until by request.

6.1.2.2.3 Colour and Formatting

The use of colour and other formatting techniques provide a number of advantages as well:

Highlight Abnormal Values. Both the group heading and the individual tests are highlighted by colour, a symbol ("!" for group headings, "v" or "^" as low or high indicator in individual test records), and the bolded font in individual test records. This is to accommodate devices without colour, or users who may have difficulties with colour perception. By making these abnormal records perceptually distinct, the display guides the physicians to the tests that require their attention. Also, their common colour and font boldness help to perceptually group these abnormal tests together despite their spatial separation.

Highlight Important Layers. A number of visual layers are created by the use of colour and font:

1. The lowest layer is the supporting information on the test, including the reference range and specimen collection date, the note indicator and patient demographics. The background colour of these areas is white (Figure 17).

1. The lowest layer is the supporting information on the test, including the reference range and specimen collection date, the note indicator and patient demographics. The background colour of these areas is white (Figure 17).

2. The next layer displays the main part of the report: test names and their corresponding values. Two sub-layers are created for the normal and the abnormal tests.

    a. The first sub-layer displays test names and values which are within the nominal ranges and is created by the use of non-bolded fonts and an unsaturated pale-yellow background colour, which is similar to the white background of its immediate lower layer (Figure 18).

    b. The second sub-layer displays abnormal test names and their values. The background colour is a more salient pale red, reinforced by the bolded text to create a slightly higher layer (Figure 19).

3. The third layer displays the group headings. By using more saturated colours, the group headings appear to be yet another layer above the test name and test value area. Theoretically, the abnormal test area is more important than the group headings in a laboratory report, and should be placed a layer above. Doing so will create a perceptual competition between the mental and visual hierarchical organization scheme, where the tests should be grouped *under* the headings, and the effect created by colour and fonts, where the tests are perceived as being *above* the headings. In fact, the mental model of the hierarchical grouping and the familiar spatial layout of hierarchical structure are powerful enough to destroy the layering effect. As a result, the group-heading layer was put above the abnormal-test layer, thus visually reinforcing the hierarchical organizational scheme to maintain a harmonious perception (Figure 20).

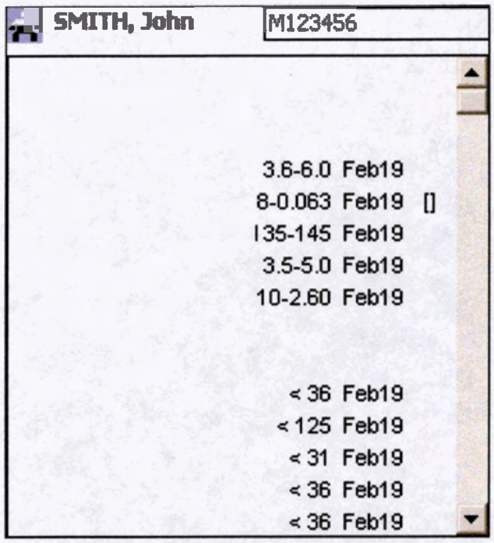


Figure 17. Layer 1 of the display.

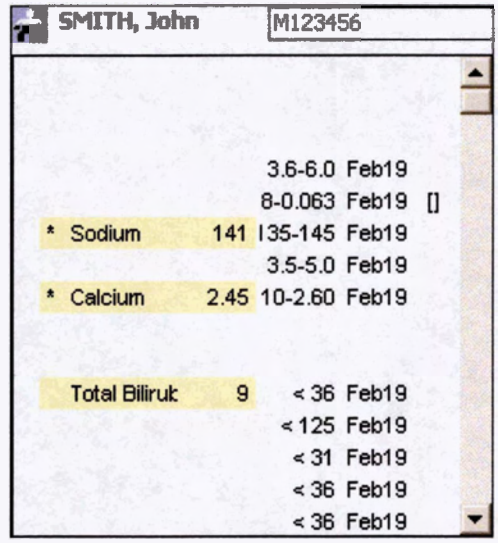


Figure 18. Layers 1 and 2a of the display

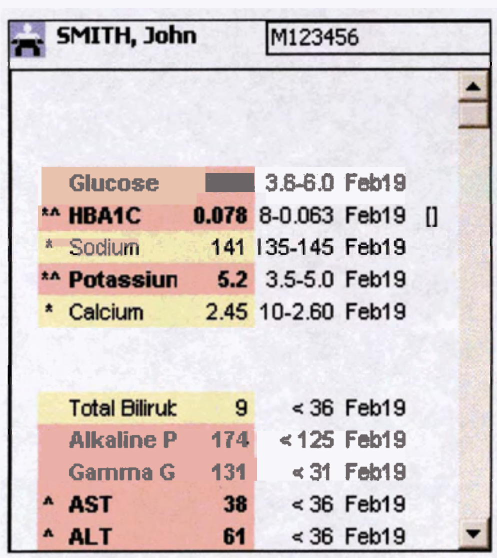


Figure 19. Layers 1, 2a and 2b of the display.

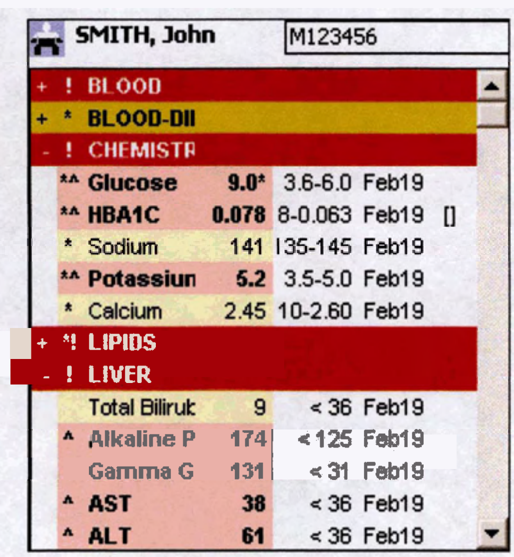
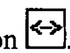


Figure 20. All three layers of the display

Layering helps direct the user's attention to important information on the display, and to reduce visual clutter despite the crowded screen. This is especially effective in reducing visual cluttering in the linear display (Figure 14). In the hierarchical case (Figure 13), visual layering helps to reinforce the structure even though the group names are aligned with the test names.

6.1.2.2.4 The Two-Views

Despite the potential benefits of the hierarchical + elision presentation technique, there are occasions where a simple linear alphabetic list will be useful. Experienced physicians working in their expert-domain usually develop elaborate schemas before reaching the laboratory report reading stage. In other words, they are very specific in their target tests of interest. Imposing an organizational scheme that may or may not fit their schemas may impede their decision-making process. In this case, it may be easier to list all the tests ordered alphabetically to allow for fast location of the pertinent results. The second situation where a linear alphabetical result may be of use is when the physician is looking for a specific test, but is unfamiliar with the organizational scheme. In that case, a simple linear list will better support the search. For these reasons, a linear list is provided along with the hierarchical one, and switching between the two views is accomplished by tapping the button .

The button  allows switching from the detailed test view (hierarchical or linear) and the pictorial overview.

Originally, the fisheye sampling method (Furnas, 1999) was considered to allow selective viewing of the tests in the linear view (§5.4.3.1). However, since the length of the test list is not excessive (up to 30 tests per report), the data processing cost is likely to outweigh the benefit of using the fisheye sampling method.

6.1.3 Other Features

To allow more seamless workflow, the button  reorders the current test or test category in case of suspicion of errors. A visual confirmation is provided in the form of a dialog box (Figure 21).

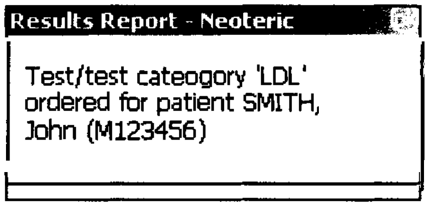


Figure 21. Dialog Box indicating a repeated Test Order.

6.2 Evaluation: Observation User Study

An observational user study was conducted to evaluate the initial design and to solicit user responses to direct further design and development.

6.2.1 Experiment design

6.2.1.1 Participants

Eight physicians participated in the study, with six from Family Medicine and two from Psychiatry. Three of the eight participants were female and five, male. Five out of six of the family physician participants were experienced clinicians with more than ten years of clinical practice, and both psychiatry participants were experienced practitioners.

6.2.1.2 Materials and Task

The iPAQ hp 3850 Pocket PC PDA[+] was used to run the software. The clinical case consists of a scenario with chief complaints, present medical history, and relevant physical examination

---

[+] Details of the hp3850 device:

Operating System: Microsoft Pocket PC 2002
Processor: 206 MHz Intel Strong ARM 32-bit RISC Processor
Display Type: Color reflective thin film transistor (TFT) LCD, 64K colors

Resolution: 240 x 320
Pixel Pitch: .24 mm
Viewable Image Size: 2.26 x 3.02 inches
RAM: 64 MB
ROM: 32 MB

findings were provided on paper (Appendix B) as clinical context to the laboratory report used in the software. The clinical case was taken from a website for training family medicine interns.

6.2.1.3 Protocol

Each session was conducted separately with the physician, and on average took 30-45 minutes. The sessions were not recorded electronically, but verbal comments by the participants and areas of confusion were noted. The protocol of the study was as follows.

1. Participants were asked to study a clinical case (Appendix B) with the goal of diagnosis. The "laboratory" section of the clinical case was presented by the software. The participants were encouraged to explore the software even after they had come to a diagnosis.

2. Participants were asked to fill out a questionnaire (Appendix B) to solicit their reactions and impressions of the software.

3. A brief 5-minute interview was conducted to cover concerns not addressed by the questionnaire.

6.2.2 Study Results

Two types of questions were included in the questionnaire: eight Likert-like multi-point scaled questions, and two open-ended questions. Responses to all questions were elicited using a 7-point continuous scale, with 1 being "strongly disagree" and 7 being "strongly agree". The basic findings of the scaled questions are:

1. The interface was felt to be very easy to use and understand, but may take a little learning. Some found using a handheld device a bit difficult at first, and felt that using the device would require adaptations.

2. The system was believed to be helpful and contained most of the expected functions.

3. The system was fun to use.

More specifically, questionnaire results expressed in terms of mode (highest frequency score) and median (50th percentile) are shown in Table 4.

Question		Mode	Median
1	It is easy to find the laboratory test result I need using this system.	7	6.5
2	It is easy to understand the meaning of the results displayed using this system.	5	5.5
3	It was easy to learn to use this system.	4	5.2
4	The organization of information on the screens is clear.	5	5
5	The information complements the way I think about the case presented in the study.	6	6
6	This system has all the functions and capabilities I expect it to have.	6	6
7	Overall, I believe this system will be more helpful to me in my patient management than the paper reports.	6	4.5
8	Overall, I like using this system.	6	6

Table 4. Scaled questionnaire result of the observational user study.

Overall, the questionnaire results show that the prototype was very well received by the participants, given that most of the questions were answered favourably (with mode and median values above 5). The responses to question 3 were partially clouded by the physicians' lack of experience with PDAs, and many believed learning to use the device was part of the learning required to use the system. As for question 7, the responses were quite dispersed, ranging from 2 (two responses) to 6 (three responses). The low scores were from the experienced family physicians, who were also experienced report readers. It is therefore not surprising that they would prefer the paper reports to the prototype.

The open-ended questions were used to solicit the participants' opinions regarding the most positive and negative aspects of the system. The participant counts for the top-three comments are included in brackets.

For the reported *positive* aspect(s) of the system:

- Good use of colour (4)
- Speedy/easy access to results (3)
- Clean and organized interface (3)
- Portability (3)
- Simple to use (2)
- Interface to LIS
- Collapsible view handy to get a tabular overview on small screen
- Normal reference values very useful
- Note editing very helpful

For the reported *negative* aspect(s) of the system:

(i) Regarding the designed display

- Grouping differs from standard lab reports (2)
- Cluttered initial screen
- Alphabetical listing useful, but non-standard
- The symbol "<->" is not easy to understand

(ii) Regarding missed functionality

- Need to see trend data (3)
- Could use an alert feature
- Lack of printouts
- Need more complete audit trail
- Concerns about confidientiality of results

(iii) Regarding data presented

- Results not in Candian units (3)
- Glucose test should not be assigned to the liver

(iv) Regarding device capability

- Concerns about resolution on non-coloured screens
- Lighting is not ideal

(v) Other concerns

- May need EPR[+] in place first

In short, most physicians believed they could and would use the system. However, there was a split on the use of anatomical/physiological organizational scheme. Experienced laboratory report readers (experienced family physicians) felt uncomfortable about its deviation from most laboratory report organizational schemes, while inexperienced laboratory report readers (new doctors and doctors in disciplines with less exposure to laboratory reports) found the body overview and the organizational scheme corresponded well with their understanding of the clinical case and overall, their understanding of the underlying medicine. One of experienced physicians believed that while the body overview may not be useful to him as a physician, it would be useful for patient education. Another experienced family physicians believed it may be useful to nursing and administrative staff, even thought it may not be useful to experienced doctors.

The need to see previous results was reported to be very important by the experienced family physicians. However, since this is beyond the scope of the project, the feature will not be added at this point.

[+] EPR, or electronic patient record.

6.2.3 First Redesign: Design Version 1

The most interesting finding of this study in terms of display techniques was the high resistance for experienced physicians to adapt to the anatomical organizational scheme. Two design changes were made to reduce this resistance:

1. ***Reconciliation of Anatomical Scheme with Current Laboratory Categories.*** As commented by an experienced family physician, a large number of tests can be easily incorporated into the anatomical scheme (e.g., Kidney function test, Liver function test), while the rest may be more difficult (e.g., blood glucose level). Due to the familiarity of experienced physicians with standard laboratory display formats, the organizational scheme for the hierarchical display and the body overview were modified to better reflect the standard formats. Tests with obvious anatomical/physiological system associations (examples include Liver Function Tests, Renal Function Tests and cultures) are assigned to these systems. Tests that are considered systemic (examples include glucose and the electrolytes) are put under their current categories in standard laboratory reports (for example, chemistry). For tests that are systemic in nature, but frequently used as strong indicators of specific organ disease, put under the anatomical rather than systemic categories. Examples include Lipids (indicative of cardiovascular diseases), tumour-markers, and microbiology cultures.

   As an example, in the new design screen capture (Figure 23), Liver Function Tests (AP, GGT, ALT, AST) are associated with the liver. Despite being a systemic test, the Lipids (cholesterol and triglycerides) are grouped under the heart since they are strong indicators of cardiovascular diseases. Hematology tests (e.g., hematocrit, or HCT) are put under the "Blood" system while systemic tests (e.g., chemistry tests like glucose, haemoglobin A-1C and potassium) are grouped separately and not associated with the body picture.

   Another concern expressed by some physicians is the ordering of individual tests within a group. In the old design, tests were arranged alphabetically. However, it turns out that there is a specific ordering system used by most laboratories. This ordering scheme is also preserved in the new design.

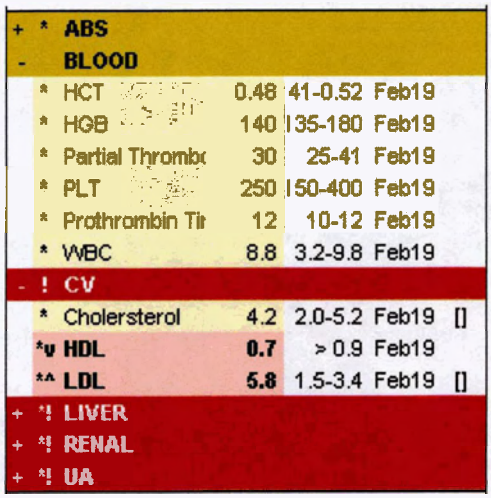


Figure 22. Original Design of the Hierarchical Display of Laboratory Tests.

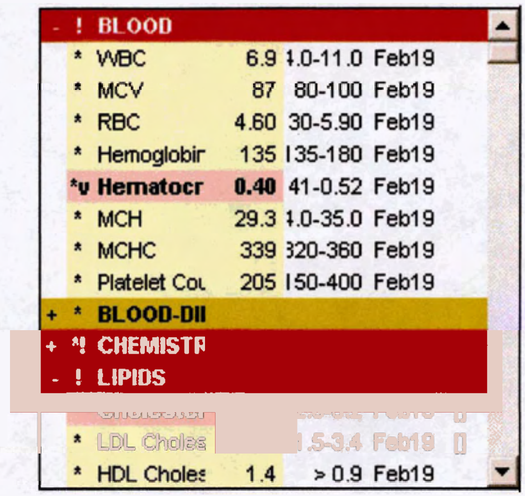


Figure 23. Design 1 of the Hierarchical Display of Laboratory Tests.

2. ***Decluttering of the Overview Screen.*** The reservations and objections raised by experienced physicians regarding the benefits of the body overview were mainly attributed to physician's familiarity with standard laboratory grouping scheme by the physicians. Nonetheless, the cluttered old overview screen was not optimal in demonstrating its potential benefits as a tool for patient status assessment (Figure 24).

   The initial body-overview screen was decluttered to avoid unnecessary visual interferences (Figure 25). Normal test values are no longer shown on the overview screen. Also, the anatomical/physiological system headings are removed since physicians are very familiar with the graphical display and the laboratory tests. On the body outline picture, only abnormal systems are highlighted, and systems with normal test results are only shown in wire-outlines.

Test names, instead of test results, are highlighted for style consistency with the hierarchical and linear lists in the main screen.

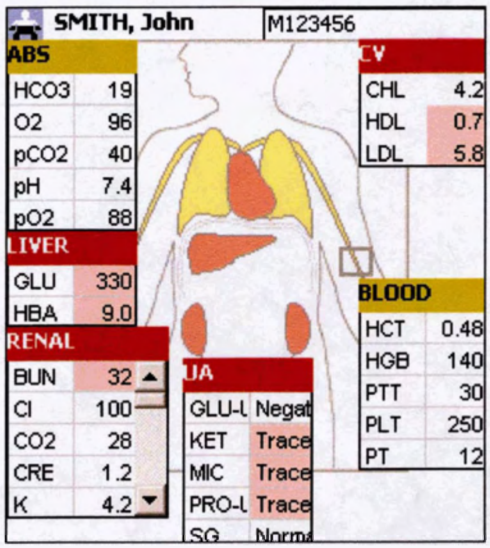


Figure 24. Original Design of the Anatomical Overview Screen.

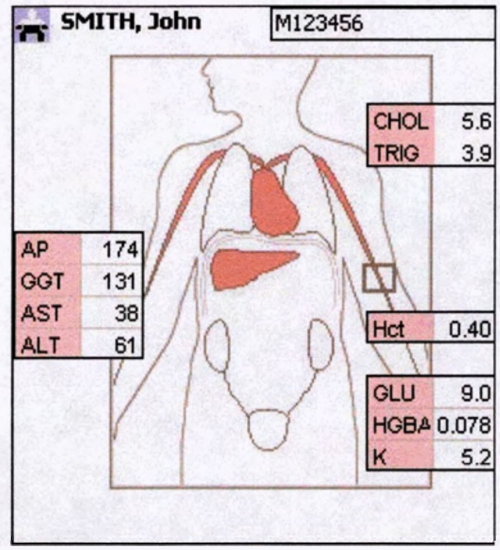


Figure 25. Design 1 of the Anatomical Overview Screen.

Based on user inputs, other refinements include:

1. ***More extensive annotations.*** Annotations are extended to all the tests, instead of those with notes from the laboratory. Physicians can time-stamp, check, and/or add a question mark to any laboratory test result. The symbol "x" is used to denote a physician's note, while "()" is for the laboratory. With both, the symbols are combined to be "(x)".

2. ***Full test name view.*** Due to the width limitation of the table columns, sometimes entries cannot be fully displayed. The new design added a feature to display the full text entry upon a single tap (Figure 27).

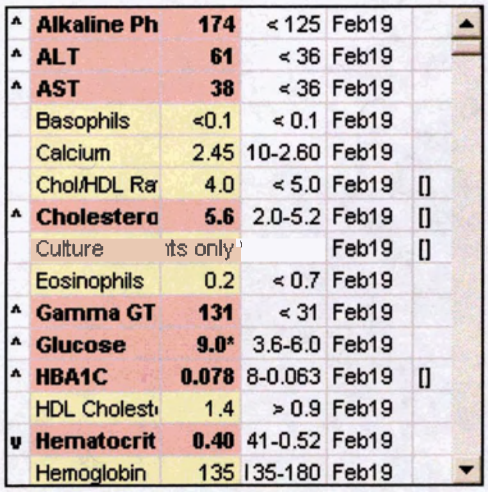


Figure 26. The original display before tapping.

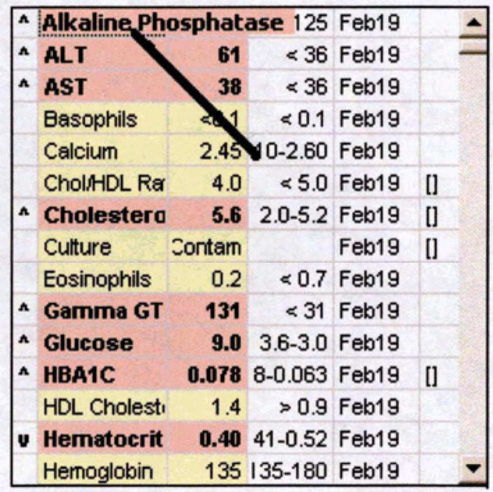


Figure 27. Elision Technique in Showing the Entire Name of a Long Test Name.

3. ***Correction of test units.*** Since the clinical case was taken from an American website, the units used were not the same as those used in the Canadian system. Some physicians found it difficult to truly evaluate the interface since they were constantly distracted by the unfamiliar result ranges. A local laboratory report was obtained at the clinic as the data for the next design.

4. ***Trend graph mock-up.*** Since all of the experienced family physicians expressed strong concerns over the lack of previous test viewing, a mock-up was included to allow better evaluation of the other aspects of the interface in the second session (Figure 28).

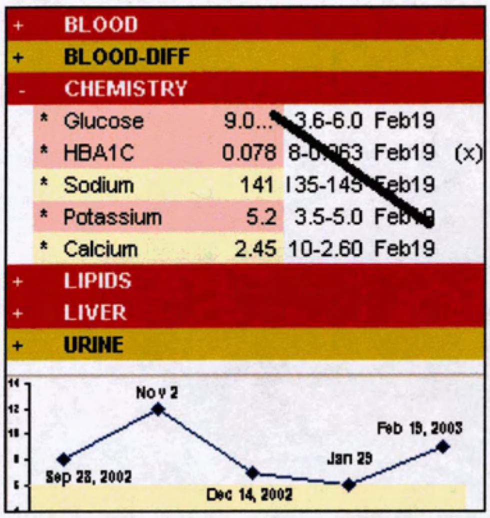


Figure 28. Graphical display of previous data (mock-up).

Previous results of the test (in this case Glucose) is shown as a chronological plot.

6.2.4 Second Evaluation

Due to the busy schedule of physicians and the favorable reviews from the inexperienced report readers, the new design was shown to the experienced report readers in the second evaluation. Also, the second sessions were less formal and consisted of verbal feedback only without a questionnaire. Four physicians, all experienced family practitioners, participated in the second evaluation. Two prototypes were shown in these session: one without the body pictorial overview, which uses the same organizational scheme for the hierarchical display based on a laboratory report obtained in the previous session; and one with the pictorial overview, and a similar organizational scheme that moves organ systems specific tests out of the general chemistry group to the corresponding organ systems.

Both organization schemes were well received by the physicians. Noting the size of the display and the number of tests that can potentially appear in the chemistry category, some physicians preferred to have the organ specific tests extracted out of the general chemistry category for more flexible viewing and filtering, as otherwise the whole screen can easily be taken up by tests under the chemistry group.

The decluttered pictorial overview was better received than the original design. While still unconvinced that the overview is essential for diagnosis and patient management, the physicians felt that the visually cleaner overview may help direct their attention to areas of interest, since all the abnormal values can be seen on a single screen.

The family physicians' reactions and comments on the pictorial overview were expected and cannot be explained solely by their being experienced practitioners and thus expected to be resistant to change, , since the two psychiatrist participants were also experienced practitioners, but still favoured the pictorial overview. The reason behind the different preference may be the difference in their clinical experience. Over the years, expert physicians must have incorporated the text-based report in their decision-making and schema formation processes. However, visually depicting the pathology on a body picture has not been supported by their current tools. It is therefore not surprising for them to find the picture non-essential in their diagnostic and patient management processes, while physicians who were less experienced with the text-based report found the picture attractive and useful. Also, studies in medical cognition revealed that expert physicians use laboratory reports to rule-in and rule-out hypotheses (§2.2). In that case, the pictorial representation is superfluous since the physicians already have an organized schema in mind. For novices, they may use the reports to development hypotheses. Having a pictorial depiction of an overview may help them organize clinical and laboratory findings in the process of diagnosis.

6.2.5 Final Design: Design Version 2

A number of design ideas came out of the second evaluation sessions.

1. ***Disease Profiles***. While the current software aimed at supporting the diagnostic process, long-term patient management is also an important aspect of medicine, especially family medicine. It would be very useful to view selected test results taken over a period of time to obtain a chronological overview of the patient from the perspective of pathology. For example, two possible disease profiles are diabetes and heart failure. Figure 29 shows the paper form of a temporal profile of diabetes, and Figure 30 shows the small screen implementation. As in the cases of the hierarchical or linear displays of test results, the abnormal values are highlighted.

		GUIDELINES		PROCEDURE						
3 TO 6 MONTHS	GLYCEMIC CONTROL	Fasting or pre-meal glucose level 4-7 (mmol/L) ; 1-2 hrs after meal 5-11 (mmol/L)		Review glucose records						
		Every 3-6 months: target <0.070		HbA_{10}						
				Hypoglycermia						
	HYPER-TENSION	Target ≥ 130/80		BP						
				BP medications						
	OTHER	Target body mass index (BMI): 20-27		Kg/m^2						
		Foot care		Lower extremity exam						
		Reinforce lifestyle counseling		Smoking, activity, diet, stress						
ANNUALLY AND/OR AS INDICATED	LIPIDS	Fasting lipid profile	Target < 2.5 mmol/L	LDL						
			target < 2.0 mmol/L	Triglycerides						
			Target < 4.0	Ratio (total chol/HDL)						
				Lipid lowering meds						
		Meter within 15% of simultaneous lab value		Fasting glucose meter/lab comparison						
	RENAL	Screen for nephropathy Type I: if > 15 yrs old & 5 yr hx, annually Type II: At diagnosis, then annually		Microalbumin screen (albumin:creatinine ratio)						
	EYES	Dilated eye exam: Type I: if > 15 yrs old & 5 yr hx, annually Type II: At diagnosis, then 1-2 yrs		Opthalmologist/optometrist for dilated eye exam						
	NEUROPATHY	Check for peripheral anesthesia		Test for loss of sensation with 10g monofilament on circled areas						
		Check for symptoms or findings such as pain, erectile dysfunction, gastrointestinal disturbance		History and physical						
	EDUCATION	People with diabetes need the support of an interdisciplinary team of health and other professionals		Diabetes/lipids education						
		Annual influenza vaccine Once-in-a-lifetime pneumococcal vaccine								
		See clinical progress notes								

Figure 29. A typical diabetic patient care flow-chart.

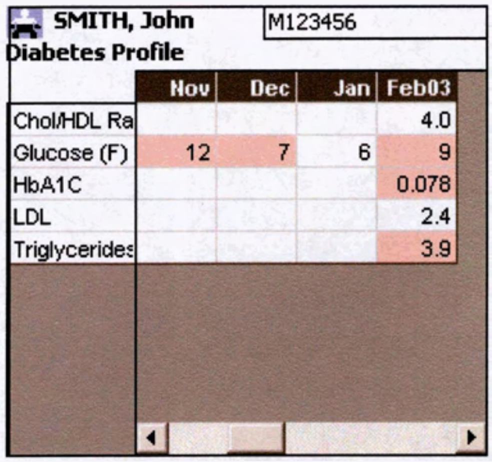
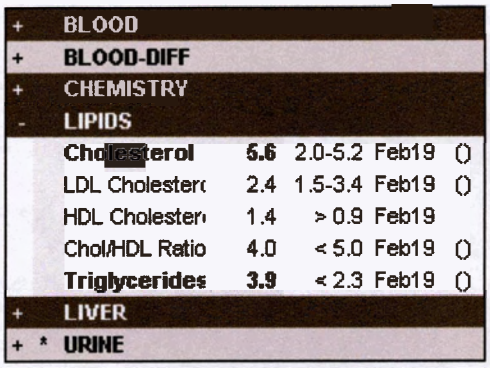


Figure 30. Diabetic Profile.

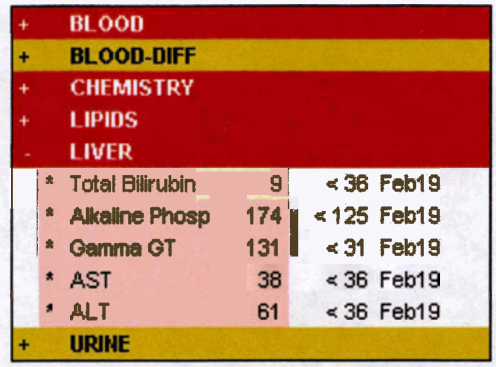
Tests that are considered to be useful in diabetes management are shown in this temporal profile. Results from previous laboratory works are incorporated and abnormal values are highlighted.

2. ***Decluttering of main screens***. The original idea of using a dual encoding for abnormal results with colour and symbol resulted in a rather cluttered screen. With a small display, the screen was found to be too busy visually. Instead, the design for the gray-scaled version was separated from that of the coloured version, where luminance and text formatting are used in the black-and-white (Figure 31, Figure 33), and luminance and hue are used for the colour (Figure 32, Figure 34). The abnormal indicators ("!", "v" and "^") were removed since the colour or bolded text was thought to be sufficient in conveying the abnormal status of the test result. According to physicians who participated in the observational study, these indicators were unnecessary for three reasons. First, a quick cross-reference to the nominal range would provide the same information. Second, once the test is found to be abnormal, it is important to determine the degree of deviation. The simplistic high and low indicators are insufficient in this regard, making the cross-referencing to the nominal ranges an essential step in the report reading process. Third, most abnormal results tend to be high or low consistently. For example, tests like cholesterol, triglycerides and glucose tend to be high when abnormal, while HDL tends to be low.



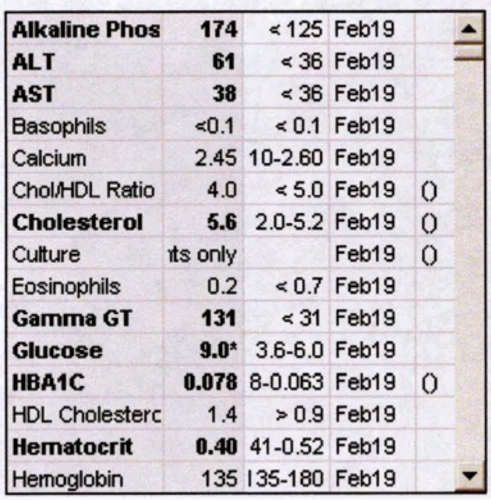
+	BLOOD				
+	BLOOD-DIFF				
+	CHEMISTRY				
-	LIPIDS				
	Cholesterol	**5.6**	2.0-5.2	Feb19	()
	LDL Choleste r(	2.4	1.5-3.4	Feb19	()
	HDL Choleste r(	1.4	> 0.9	Feb19	
	Chol/HDL Ratio	4.0	< 5.0	Feb19	()
	Triglycerides	**3.9**	< 2.3	Feb19	()
+	LIVER				
+ *	URINE				

Figure 31. Black-and-White View of the Hierarchical Test Display.



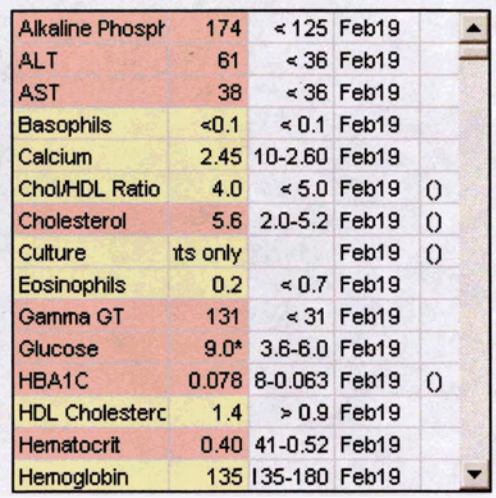
+	BLOOD			
+	BLOOD-DIFF			
+	CHEMISTRY			
+	LIPIDS			
-	LIVER			
*	Total Bilirubin	9	< 36	Feb19
*	Alkaline Phosp	174	< 125	Feb19
*	Gamma GT	131	< 31	Feb19
*	AST	38	< 36	Feb19
*	ALT	61	< 36	Feb19
+	URINE			

Figure 32. Coloured View of the Hierarchical Test Display



Alkaline Phos	**174**	< 125	Feb19	
ALT	**61**	< 36	Feb19	
AST	**38**	< 36	Feb19	
Basophils	<0.1	< 0.1	Feb19	
Calcium	2.45	10-2.60	Feb19	
Chol/HDL Ratio	4.0	< 5.0	Feb19	()
Cholesterol	**5.6**	2.0-5.2	Feb19	()
Culture	its only		Feb19	()
Eosinophils	0.2	< 0.7	Feb19	
Gamma GT	**131**	< 31	Feb19	
Glucose	**9.0***	3.6-6.0	Feb19	
HBA1C	**0.078**	8-0.063	Feb19	()
HDL Cholesterc	1.4	> 0.9	Feb19	
Hematocrit	**0.40**	41-0.52	Feb19	
Hemoglobin	135	135-180	Feb19	

Figure 33. Black-and-White View of the Linear Test Display.



Alkaline Phospl	174	< 125	Feb19	
ALT	61	< 36	Feb19	
AST	38	< 36	Feb19	
Basophils	<0.1	< 0.1	Feb19	
Calcium	2.45	10-2.60	Feb19	
Chol/HDL Ratio	4.0	< 5.0	Feb19	()
Cholesterol	5.6	2.0-5.2	Feb19	()
Culture	its only		Feb19	()
Eosinophils	0.2	< 0.7	Feb19	
Gamma GT	131	< 31	Feb19	
Glucose	9.0*	3.6-6.0	Feb19	
HBA1C	0.078	8-0.063	Feb19	()
HDL Cholesterc	1.4	> 0.9	Feb19	
Hematocrit	0.40	41-0.52	Feb19	
Hemoglobin	135	135-180	Feb19	

Figure 34. Coloured View of the Linear Test Display.

3. ***Displaying Trend Data.*** Being able to view previous test results is an important feature in laboratory report software. However, due to the different kind of visualization issues involved in time-sequence plots, the problem was considered beyond the scope of the project. Instead, a mock-up was added as a start for future development. In the mock-up, tests with previous results have the ellipsis symbol ("…") trailing the current result. In Figure 28, the patient has previous test results for the test glucose in the laboratory

information system, which is indicated by the ellipsis trailing the current result "9.0" (circled). Once tapped, a graphical display of previous results will be shown at the bottom of the display.

Despite the mixed reviews of the pictorial overview screen, it is kept in the final prototype for three reasons:

1. It may help new physicians in their development of expertise, since it provides an explicit visual overview for them to organize clinical and laboratory findings. This may help their development of problem analysis.

2. Even thought current expert report readers did not find the pictorial overview useful as physicians, they found it useful in patient and staff education. The pictorial overview may therefore be used as an education tool for the physicians.

3. The estimation of usefulness of a tool is probably based on available tools to the physicians while they were developing their diagnostic skills. Given an alternative, some future experts may incorporate the pictorial overview in their decision-making processes. This is similar to the case in radiology. Current radiologists are very well trained on visualizing pathologies using 2-dimensional slides, and may not find 3-D rendering of those slides useful. However, novices in the field may appreciate the 3-D rendering, and may eventually incorporate the tool in their work processes.

6.3 Conclusions from Prototype Design Process

Guided by the investigations studied in chapters 2 to 5, a series of design prototypes were developed using the iterative design process. The final prototype supports:

- Rapid patient status assessment with the overviews (the body diagram and the hierarchical headings), using various display techniques including rapid vision and object display;

- Diagnostic decision making for both experts and novices by providing a hierarchical and a linear display of laboratory test using various display techniques including perceptual layering and "detail-on-demand";

- Long-term patient management by providing a temporal profile of tests that can reflect the status of the disease.

A number of design assumptions were made in developing these prototypes. Examples include using the object display for rapid status assessment, and the use of hierarchical presentation technique in displaying lists. These design assumptions were evaluated by an observational user study specifically for the laboratory report application. It would be interesting to establish guidelines for these presentation techniques in a broader sense for small screens. Specifically, the questions may be:

1. Is object display effective in rapid information gathering and status assessment?

2. What are the performance costs when the schema used in a hierarchical display does not match those of the users?

3. What are the costs and benefits in using the hierarchical + elision presentation technique in displaying lists of data when compared to the conventional scroll lists?

The next chapter outlines three experimental designs that were developed to investigate these issues.

7 Quantitative Evaluation Plans of Final Design

While the observational user study performed with the target users was helpful in the overall design evaluation and design development of the prototype, specific aspects of the design have to be evaluated by formal comparative user studies. Three experiments were designed to assess the validity of a number of design choices. One of the three experiments was further developed and conducted.

7.1 Experiment 1: Patient Status Assessment

To mitigate the limitation of small screen displays, it was suggested that a pictorial presentation of data would be more effective than a tabulated alphabetical list in rapid status assessment. An experiment was designed to investigate this assumption.

7.1.1 Data

Abstract data instead of real laboratory tests will be used for the experiment to avoid biases from participants' previous experience and knowledge of laboratory tests. All items in the list will be in the form of `<random letter>-<organ system>` and coupled with a random numerical value. An example of an item on the list would be "`C-HEART 6.2`". Each list will have four categories, with five items per category, resulting in a total of 20 items per list. 30 lists will be generated for the experiment.

7.1.2 Displays

Data will be presented either in the form of a tabulated alphabetical list or pictorially using the anatomical overview. To avoid unnecessary exploration and manipulation of equipment, and to better control display times, all lists will be displayed on a standard 17-inch computer. For the tabulated alphabetical list, the size of the list will be adjusted to accommodate 20 items in 12-point Arial font. As for the pictorial display, the display size will be the same as the real device. Abnormal values will be coloured red on both displays.

7.1.3 Participants

20 participants from the university will be recruited for the user study. No special selection criteria are required other than colour-blindness.

7.1.4 Method

Two presentation techniques (tabulated alphabetical list versus pictorial) and three levels of test result abnormality (0, 50 and 100% abnormal tests) constituted the two factors in this experiment. Each participant will perform a block containing five replicates of the experiment. Each replicate consists of six status assessment tasks, one for each combination of the different factors and levels. The experiment is expected take about 20 minutes. For each task,

1. A blank screen will be displayed for 200 milliseconds;

2. A laboratory report will be briefly displayed for 500 milliseconds using either of the presentation technique;

3. The screen will be blanked. Participants will be asked to identify the abnormal organ system(s) by selecting from a list of options displayed at the end of each task. The selection time is limited to 30 seconds.

The order of the technique and abnormality level presentation will be randomized for the experiment, and each trial will use a different list. A training session will be provided to familiarize the participant with the task and the two presentation techniques.

7.1.5 Response Variable

The effectiveness of the presentation technique is measured by error rate, which is defined by the total number of incorrect selections (either as selecting an incorrect system, or as failing to select the correct system) at each trial.

7.1.6 Result Analysis

The results will be plotted as "Task Error" versus "Degree of Abnormality" for each presentation technique. Regression testing will be performed.

7.2 Experiment 2: Diagnostic Support (Schema Matching)

It is assumed that if the organizational structure of the display matches that of the physician's, the system will better support the task of decision-making.

7.2.1 Data

Abstract data will be used for the experiment. All items in the list will be in the form of <label><value>, where <label> is a random 3-letter label, and <value> is a random number. An example of an item on the list is "AEX 6.2". Each list will have five categories, with five items per category and 25 items per list. Three lists will be generated for the experiment. In addition, a list of rules will be given for the participants to answer the question. This is to mimic the rule-based decision-making logic similar to that used by physicians. One sample rule is If AEX ≥ 4.5 and BVC ≤ 3 then CONDITION 1 is TRUE. Three sets of rules will be generated for the experiment.

7.2.2 Displays

Data will be presented in a hierarchical display on a standard 17-inch computer monitor. The size of the display will be adjusted to accommodate 25 items in 12-point Arial font. Along with the list, a corresponding question will also be displayed.

7.2.3 Participants

20 participants from the university will be recruited for the user study. No special selection criteria are required.

7.2.4 Method

There are three stages for this experiment: No predefined schema, matched schema, and mis-matched schema. Due to learning effects, each participant can only perform a single block consisting of three tasks. This experiment is expected to take 5 minutes.

A list of 25 items will be displayed in a hierarchical organizational scheme based on the second character in the test label. The participant is asked to answer a number of rule-based logic questions based on the data and the rules given. They are encouraged to answer the questions as accurately and as quickly as possible.

1. ***List 1—no schema case***: The first question tested using the first list is the "no preconceived schema case", since the participants will not be expecting a specific organization scheme. The participant will answer four more questions to ensure they acquire the scheme used in the first list, but these questions will not be used in the analysis due to learning effects.

2. ***List 2—correct schema case.*** The participants will answer one question for this case.

3. ***List 3—incorrect schema case.*** The list will be organized randomly. That is, it will not be based on the prefix of the test names. The participants will answer one question for this case.

The ordering of the lists will be randomized. No training session is required for this experiment.

7.2.5 Response Variable

The effect of schema matching will be determined by task performance time, which is defined as from the start of the display (list and question) to the time the participant selects an answer.

7.2.6 Result Analysis

Using the "no schema case" as the baseline, the calibrated "correct schema case" results will be compared to those in the "incorrect schema case" to determine the effect of schema matching in rule-based decision-making.

7.3 Experiment 3: Diagnostic Support (Grouping)

Medical decision-making usually requires more than one piece of data. If the data could not be displayed on the same screen, the user would need to remember the information while navigating to the next piece of information. Hierarchical display with elision is argued to be a more effective presentation technique in these situations. First, categorical grouping physically groups similar data to avoid "cross-screen" referencing. Second, even if the data are grouped in different categories, the elision technique allows hiding of data to lessen the amount of "cross-screen" referencing that would otherwise be unavoidable. This experiment is designed to test the effectiveness of hierarchical display with elision as a presentation technique when compared to vertical scrolling, a common technique in displaying long lists.

7.3.1 Data

Abstract data will be used for the experiment. All items in the list will be in the form of `<random letters>` and coupled with a random numerical value. An example of an item on the list is "`AEX 6.2`". Each list will have five categories, with seven items per category totalling 35 items per list. 25 (24 trials + 1 training) lists will be generated for the experiment. In addition, a list of rules will be given to the participants, which will be required in answering the question. This is to mimics the rule-based logic similar to that used by physicians. One sample rule is `If AEX ≥ 4.5 and BTV ≤ 3 then CONDITION 1 is TRUE`. A single of set of 12 rules will be generated for the experiment to avoid learning effects.

7.3.2 Displays

Data will be presented either in a hierarchical display or as an alphabetical list, and will be displayed on a standard 17-inch monitor for better control and smoother task flow. The actual displays will be the same size as the target PDA. For the hierarchical display, the organizational scheme will be based on the first letter to ensure familiarity and to avoid possible learning effects.

7.3.3 Participants

20 participants from the university will be recruited for the user study. No special selection criteria are required.

7.3.4 Method

This experiment will consider the effect of data distribution on the two displays. Four scenarios of physical separation of the data required in making the decision will be investigated.

1. The data required are visible on the same screen for both presentation techniques;
2. The data required cannot be displayed on the same screen for the alphabetical list, but are under the same category for the hierarchical display;
3. The data required cannot be displayed on the same screen for the alphabetical list, but can be visible on the hierarchical display with data hiding using elision;
4. The data required cannot be displayed on the same screen for either technique.

Each participant will perform the same task for each of the presentation techniques. For each task, a list of 35 items (or 5 x 7 in the hierarchical display) will be displayed. In addition, a set of rules will be displayed on a computer monitor and will remain visible for the duration of the task. When ready, the participant will answer a list of 24 (4 scenarios x 3 trials x 2 presentation techniques) questions, displayed one at a time on the same monitor. These questions will be rule-based, and will require three pieces of information with a physically distributed as one of the four scenarios described above. The participants will be encouraged to be as accurate as possible in answering these questions.

The order of the questions, the lists and the presentation technique will be randomized. Participants will be trained with a training task using both display techniques. Training session will be provided to familiarize participants with the two presentation techniques, the task, and the rules. This experiment is expected to take 30 minutes.

7.3.5 Response Variable

The effect of schema matching will be determined by task performance time, which is defined as from the start of the question display to the time the participant selects an answer.

7.3.6 Result Analysis

The results will be plotted as "Performance Time" versus "Data Distribution" for each presentation technique. Pair-wise comparisons will be done when necessary.

7.4 Conclusion from the Quantitative Evaluation Plans

Since the main purpose of the interface is for diagnostic and decision-making support, Experiment 3: Diagnostic Support (Group) was chosen for further development and the final design was conducted. The next chapter provides the theoretical background of grouping, the testing protocol evolution, and the result and analysis of the user study.

8 The Hierarchical Presentation Technique

This chapter describes the comparative user study performed to study in detail one aspect of the prototype: the use of the hierarchical + elision presentation technique in displaying list data. The chapter begins by providing background information about guidelines in using hierarchical presentation techniques in interface design, and continues with details of the user study, including its design evolution, final protocol, results and analysis.

8.1 Hierarchical Organization

Organizing a list of data using hierarchical grouping technique is a well-known visual interface design guideline. According to Mullet and Sano (1995, p. 94):

> By grouping similar elements together, the designer helps the user deal with a complex information display by reducing it to a manageable number of units. Higher level structures orient the user and help them establish a plan for moving the attention to some interesting portion of the display for a more detailed reading.

According to Furnas (1999), hierarchical grouping has better navigability than simple linear structures in terms of two quantitative measurements: efficient view traversability (EVT) and view navigability (VN). Further discussion of these two measurements can be found in §5.4.3.1.

While this may be an obvious recommendation for desktop applications, the situation is slightly different on small screens, where screen space is severely limited. To promote visual searching based on the Gestalt law of proximity, items in the group should be indented to visually separate them from the heading. This reduces the total number of columns displayable horizontally without scrolling. Vertically, an extra row is required for category information to display hierarchically grouped list items, thus reducing the number of items displayable vertically without scrolling. Increasing the number of levels in the hierarchy, or depth, is therefore increasingly expensive in terms of screen space, as demonstrated in Figure 35.

The number of items per group, or breadth, is also an important consideration. If there are too many items in each group, the headings of the groups may not be displayable on a single screen, thus obscuring the organizational scheme and making visual item searching difficult.

Group 1	Item 1
Item 1	Item 2
Item 2	Item 3
Item 3	Item 4
Group 2	Item 5
Item 4	Item 6
Item 5	Item 7
Item 6	Item 8
Group 3	Item 9
Item 7	
Item 8	
Item 9	

Figure 35. Displaying the Same List with The Hierarchical and the Linear Presentation Technique.

In order to minimize the adverse effects of the hierarchical display technique on small screens, the prototype design limits the structure depth to two levels. For example, current paper-based laboratory reports usually have the following grouping scheme:

Chemistry>Liver Function Test>Bilirubin

Which is a three-leveled structure. This can be reduced to a two-leveled structure as:

Liver Function Test > Bilirubin

since clinically, the group "Chemistry" doesn't provide additional information to the group "Liver Function Test".

To reduce the adverse effect of potentially large breadth, the elision presentation technique was incorporated into the hierarchical display in which users can open and close each group to view or hide its items to preserve both the global and detailed views. Since items within groups can be hidden, they may be difficult to find especially if the user is not familiar with the organizational scheme. This is similar to the case in menu-selection where menu items are initially hidden, until the user selects the menu to reveal the items. In fact, the problem associated with hidden menu items is considered to be so severe that Nielson (1993, p. 55) gave the following advice:

> The best advice is obviously to avoid hierarchical menus since they hide options from the user and require the introduction of an extra set of interaction techniques for navigating the hierarchy. Therefore, it is often better to overload a nonhierarchical menu slightly than to split it into a hierarchy.

In face of these conflicting recommendations, a comparative user study was conducted to estimate the potential costs and benefits of using hierarchical presentation technique when compared to a flat linear display for a task that mimics physician's laboratory report reading.

8.2 Comparative User Study

8.2.1 Study Design Evolution

The main goal of the study is to investigate the costs and benefits of using a hierarchical + elision presentation technique over a more conventional linear one in the context of laboratory report displays. Ideally, the participants should be physicians. Due to the difficulty in recruiting physicians for the test, the task should mimic physician's laboratory report reading task without the requirement of their domain expertise.

When using laboratory reports for diagnostic decision-making, physicians generally develop a schema while interpreting the report. In other words, they tend to look for a list of results, determine their normality, and fit that into one or more schema of pathology in order to rule-in or rule-out hypotheses.

The study design underwent five iterations to develop a study task that:

- Requires evaluation based on rule-based logic statements. This is to mimic the task where the physician determines the status of a particular test result (elevated, depressed or normal)
- Involves more than one evaluation. Since very few diseases can be diagnosed by a single laboratory test, a number of evaluations are likely required.

The first design used a Number-Letter-Number combination as the name of the data point since it is similar to the familiar Canadian postal code structurally, and by using the first character as the grouping scheme, this naming convention should ensure a clear organization scheme. Each data point had a numerical value. The task was to determine the logic state of a single statement in the form of

```
<data 1> {>, <} <value 1> XOR <data 2> {>,<} <value 2>
```

The reason for using the XOR logic operation was to ensure the need to evaluate both statements before arriving at the overall logic state. An example of the rule can be seen in Figure 36.

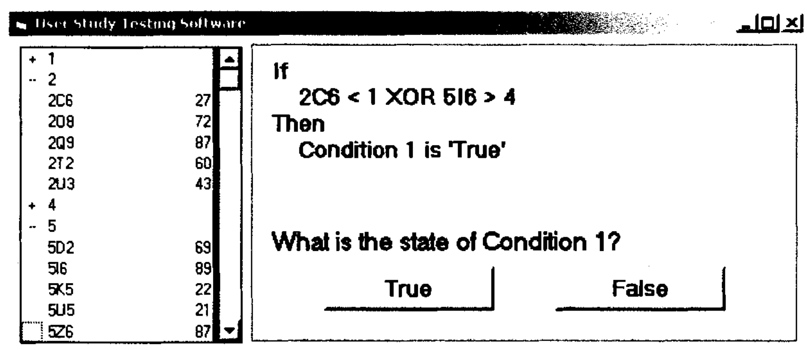


Figure 36. Design 1 of the User Study Testing Software.

In this trial, the data points are displayed using the hierarchical presentation technique, and are grouped into five categories based on the first character. The "+" and "-" indicate the state of the group as being "close" or "open".

The first design had a number of problems:

- The task exhibited a floor effect, as judged by the task times and feedback from the pilot participants. The average task time was about 15 seconds, which was probably the minimum amount of time required to determine the logic states of the two clauses.

- It was more difficult physically to click the "False" than the "True" button since the mouse was usually in the data table area.

- The task did not test the effect of data distribution, since the comparison was made between an individual data point and a given value, but not between data points.

In view of these problems, design 2 was produced:

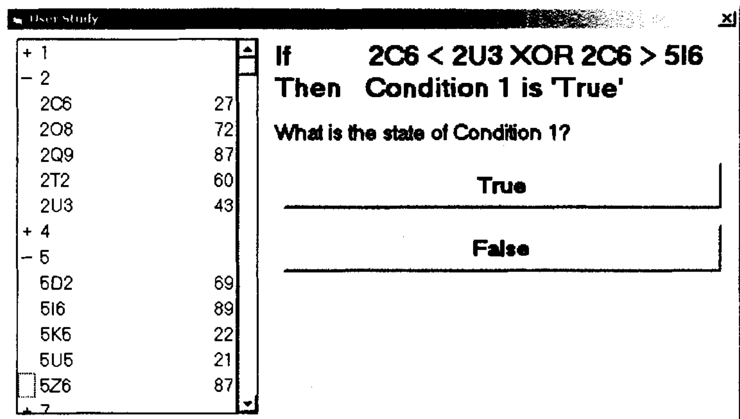


Figure 37. Design 2 of the User Study Testing Software.

In design 2, the task involved two pair-wise comparisons between data points, and the two selection buttons were equal distance from the data table. Despite the need to look up twice as many data values as design 1, the task difficulty did not increase substantially as indicated by the average task performance time of the pilot participants. Also, the XOR operation was found to be very unnatural even with a reference sheet. The pilot test participants needed to constantly refer to the logic table thus unnecessarily complicating the task. Design 3 should therefore replace the XOR logic operation. In order to ensure both statements were evaluated, XOR could not be replaced by AND, or OR. Instead, the new task required counting the total number of true statements. Also, data items in the display panel were indented to provide a better visual layout.

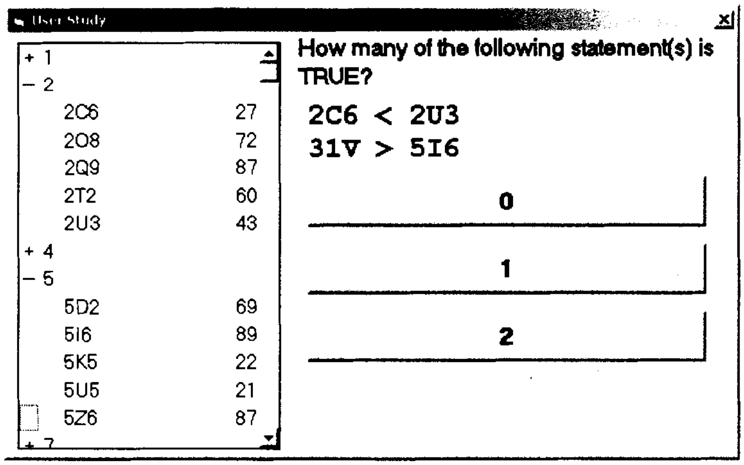


Figure 38. Design 3 of the User Study Testing Software.

A common problem in these designs was that the task was still too simple to expose potential differences between the two displays. To increase the difficulty of the task, a three data point comparison was used in Design 4.

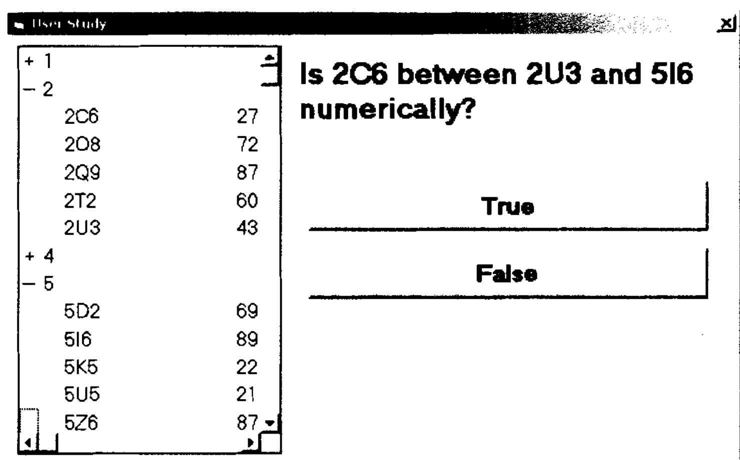


Figure 39. Design 4 of the User Study Testing Software.

While design 4 may have accomplished the desired task difficulty, the task deviated significantly from the original laboratory report-reading task. Also, some participants found the number used in the data name interfered with the data value. In addition, since the same set of data and logic statements would be used for both the hierarchical and linear displays for different participants, a data set that allowed more flexible data separations in both display techniques was required. By using the number-letter-number design with the first number as the group name, spatially close data points in the hierarchical display would also be spatially close in the linear display. A different organization/naming scheme was therefore needed.

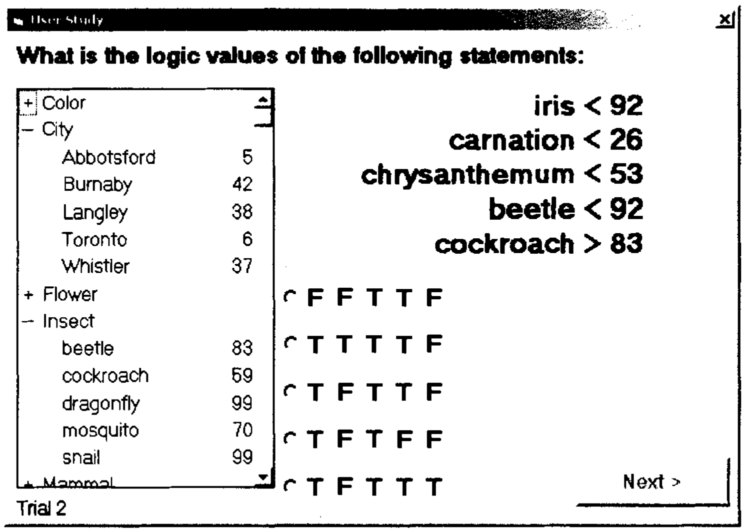


Figure 40. Design 5 (Final Design) of the Testing Software.

In design 5, the participant was required to determine the logic states of the five statements on the upper-right corner of the screen by locating the required data value and comparing to the specific value in those statements, resulting in a list of five logic states. The participant would then need to match that with the five options available in the lower half of the display, where `T` means "`True`", and `F` means "`False`". In addition, the set of answers given in the task can be used as schemas, similar to those developed by physicians before viewing laboratory reports. This design was chosen since it was believed to be a better abstraction of physician's report-reading task.

The validity of Design 5 was confirmed by an Emergency Physician, who commented, "Yes, indeed, this [the user task] simulates closely what our brain[s] do[es] when interpreting lab[oratory] values".

One drawback of using Design 5 is the stress on short-term memory. Physicians are trained to evaluate a number of hypotheses at the same time, and can fluidly switch from hypothesis to hypothesis based on results in the laboratory reports. However, since the participants here are unlikely to have developed this skill, the task, despite being similar to that of the physicians' laboratory reading task, may be more difficult for the participants than for the physicians.

Another major change in going from Design 4 to 5 is the use of real life data points. This was chosen to avoid the confusion between the data label and the values, and to allow for a more flexible data distribution in both display techniques. For example, it would be possible to request a set of data that was widely distributed in the linear display, but narrowly distributed in the linear view, since the group headings no longer determined the item names using the first five items in the alphabetically sorted linear list

8.2.2 User Study Details

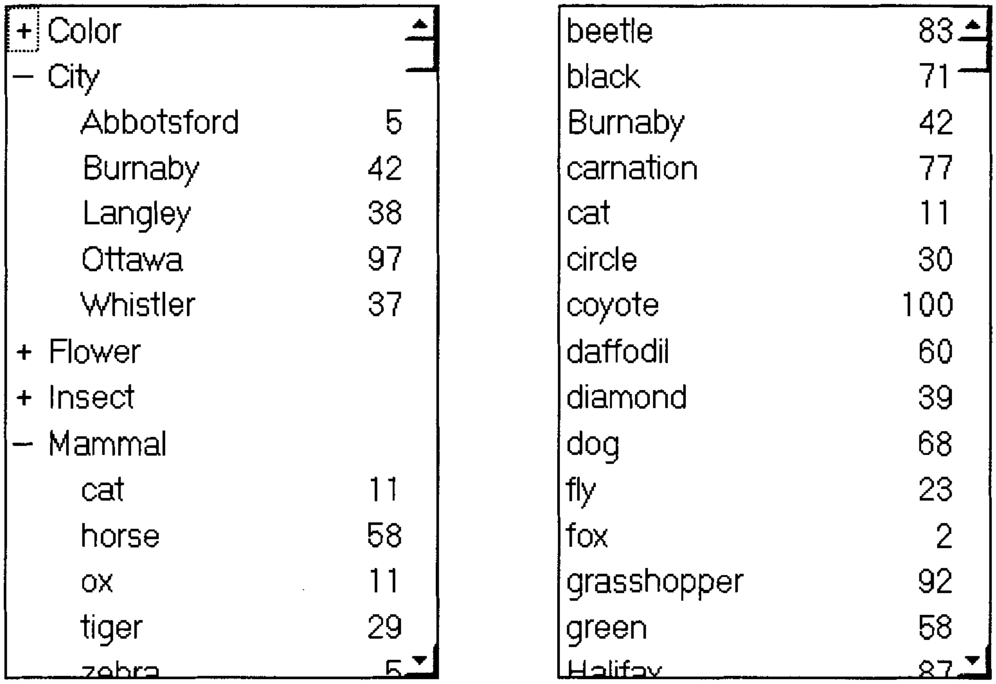
8.2.2.1 Participants

Fourteen participants (10 males, 4 females) from the university were recruited for the user study using mass emails. 10 participants were from the School of Computing Science, two from Interactive Arts, one from Information Technology, and one from Engineering Science. The mean age of the participants was 27.7 years (median = 23.5; SD = 8.3). All participants were right-handed.

8.2.2.2 Display

The testing software was displayed using the same laptop with a resolution of 1280x768, in which the data display area was 4 cm wide by 6 cm high.

Examples of the two presentation techniques are given in Figure 41.



Hierarchical Display **Linear Display**

Figure 41. Examples of hierarchical and a linear data display used in the study.

15 data points were displayed using either the hierarchical or the linear display. In the hierarchical display, data points were grouped according to the category type, with five points per category, selected from a list of ten possible candidates. There were six categories in this study: *Color*, *City*, *Flower*, *Insect*, *Mammal* and *Shape*:

Category	Member Items
Color	black, blue, green. magenta, orange, pink, purple, red, white, yellow
City	Abbotsford, Burnaby, Halifax, Kelowna, Langley, Ottawa, Surrey, Toronto, Vancouver, Whistler
Flower	buttercup, carnation, chrysanthemum, daffodil, iris, orchid, primrose, rose, sunflower, tulip
Insect	ant, beetle, cockroach, dragonfly, grasshopper, fly, mosquito, snail, spider, whitefly
Mammal	cat, coyote, dog, fox, horse, ox, pig, sheep, tiger, zebra
Shape	circle, diamond, hexagon, pentagon, rectangle, rhombus, octagon, parallelogram, square, triangle

Table 5. Categories used in user study.

Data points were arranged alphabetically within each group. Groups could be expanded or collapsed by clicking the "+" /"-" sign or the group title. For the linear display, data points were arranged alphabetically using the entire data name. For both displays, scroll bars were used for scrolling up and down the lists.

8.2.3 Experimental Design

Experiment 1:
The Effect of Data Separation and Sorting using the Linear Presentation Technique

A 2x2 within-subject, randomized complete block design was used, with the first factor being data separation with two levels (small and large) and the second being sorting, also with two levels (sorted and unsorted). Data separation is defined by the ability to view all the five data points specified in the rules within the same screen (small), or over more than a single screen (large); alphabetically sorted, or unsorted, as shown in Figure 42.

magenta < 89	pentagon < 82
octagon > 6	hexagon > 97
orange > 35	horse < 15
orchid > 56	cat < 55
Ottawa < 91	ox > 58

Figure 42. The sorting factor.

The data points required by the rules in the experiment can be either sorted (left hand figure), or unsorted (right hand figure).

The participant effect was considered as a nuisance factor and an attempt was made to eliminate this effect by blocking. Each block consisted of four replicates, and each replicate consisted of four trials, one for each of the four factor combinations. The order of presentation of the trials was random, and different data sets and logic rules were used for each trial. The performance measure was task time. Accuracy data was also collected.

Experiment 2:
The Effect of Category Sorting using the Hierarchical Presentation Technique

A single factor, within-subject, randomized complete block design was used, with the factor being category sorting with two levels (sorted or unsorted) and block by participant. The data points requested in the five logic rules were taken from two adjacent categories in the hierarchical display, and may be sorted by category, or unsorted.

Each block consisted of two replicates, and each replicate consisted of two trials for each of the two levels. The order of level presentation of the trials was random, and each trial used a different data set and logic rule. The performance measure was task time. Accuracy data was also collected.

Experiment 3:
The Effect of Data Category Distribution using the Hierarchical Presentation Technique

A single factor, within-subject, randomized complete block design was used, with the factor being data category distribution with five levels: all five data points requested in the logic rules were contained within one, two, three, four or five categories of the data displayed hierarchically. The data needed to evaluate the logic rules were sorted for the single group display, both sorted and unsorted for two-grouped displays, and unsorted for the rest. The experiment blocked by participant.

Each block of the experiment consisted of two replicates, and each replicate consisted of eight trials: two trials for groups with category separation of one to three, and one trial for groups with category separation of four and five. The order of the groups was randomly assigned with a different data set and logic rules. The performance measure was task time. Accuracy data was also collected.

8.2.4 Procedure

All trials (5 practice, 32 actual) were displayed in the format shown in Figure 40. Data were displayed using either the linear or the hierarchical presentation technique randomized using the software function `Rand()`. Each data point consisted of a name (e.g., cat) with an associated value (e.g., 11). Participants were instructed to evaluate five logic statements (e.g., `Cat > 12`) based on the data set displayed. To evaluate the statement, the participants needed to find the data point specified among the data displayed, and determined if the statement is `TRUE`, or `FALSE` based on the data value associated with the data point requested in the rules.

Each task had a set of five such statements based on a different data set. The goal was to determine the logic states of all of the statements, and match it to one of the five choices. The choices are in the form of "T T F T F", and the selection differed from trial to trial. Participants were instructed to evaluate the logic statements sequentially from top to bottom.

The study procedure was as follows:

1. *Practice session*, with five questions.
2. *Actual study session*, with 32 questions. The participants were advised to take breaks in between questions if required.
3. *Questionnaire,* with nine questions to solicit subjective impressions about the presentation techniques.

8.2.5 Results and Analysis

Among the fourteen participants, three demonstrated evidence of guessing as indicated by their reaction times. In performing the trials, some participants prematurely selected an option, but continued to examine the data. If their initial guess was correct, some may not reselect the option already chosen, thus resulting in abnormally short task completion time for that trial. As a general rule, it is impossible to complete a trial within 10 seconds, and result sets containing such abnormally short task time were discarded. In short, the data presented were based on results provided by eleven participants.

Time Data Analysis

Analysis of variance (ANOVA) was chosen to be the statistical tool in analyzing time data. Since Experiment 1 is a two-factor, randomized complete block design, the ANOVA model for the data is as follows,

$$y_{ijkl} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \varepsilon_{ijkl} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,c \\ l = 1,2,\ldots,n \end{cases}$$

where y_{ijk} is the lth observation in the ith and the jth treatment group for the kth participant,

μ is the overall mean of the general population,

τ_i is the effect of the ith treatment,

β_j is the effect of the jth treatment,

$(\tau\beta)_{ij}$ is the effect of the interaction between the ith and the jth treatment,

δ_k is the effect of the kth participant, and

ε_{ijk} is the random error in the lth observation in the ith and the jth treatment group for the kth block.

Similarly, the model for experiments 2 and 3 is,

$$y_{ijk} = \mu + \tau_i + \delta_j + \varepsilon_{ij} \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,b \\ k = 1,2,\ldots,n \end{cases}$$

where y_{ij} is the *k*th observation in the *i*th treatment group for the *j*th participant,

μ is the overall mean of the general population,

τ_i is the effect of the *i*th treatment,

δ_k is the effect of the *j*th participant, and

ε_{ij} is the random error in the *k*th observation in the *i*th treatment group for the *j*th participant.

These models assume negligible interaction between blocks (i.e., participants) and treatments, or when such interactions do exist, they are inseparable from the error components (Montgomery, p. 207).

Before using ANOVA in the data analysis, residual analysis was performed to test the assumptions of error distribution normality, error serial independence, and homoscedasticity of the ANOVA model. The assumption of normality was tested with normality plots. In a number of cases (e.g., the largely separated data sets in Experiment 1, Figure 43), the datasets were found to skew towards the right. As reported by Wilcox (1997, ch.1; 1998), experimental power is substantially lower in cases of skewness, heavy-tailed distributions and outliers. However, with the large number of samples (e.g., $n = 176$ for the datasets in Experiment 1), the effect of the skewness may be insignificant.

Plots of residuals in time sequence were used to detect any signs of serial dependence. Levene's test was used to test for homoscedasticity. Both of these criteria were satisfied.

Accuracy Data Analysis

Unlike the time data, the accuracy result for each trial is either 1 (correct), or 0 (incorrect). In this case, using a parametric procedure (e.g., ANOVA) would be inappropriate, since the assumption of normality obviously fails in this case. Instead two non-parametric procedures were used. For

Experiment 1, the two-factor Friedman's test was used and for experiments 2 and 3, the one-factor Kruskal-Wallis test was used.

Overall Results

The average task completion time across all experiments was 36.3 seconds (*SD* = 14.5 s; median = 34 s). All participants were at least 75% accurate in their trials (mean = 0.86; *SD* = 0.35; median = 1).

Experiment 1:
The Effect of Data Separation and Sorting using the Linear Presentation Technique

1. Time Results

The data distributions of the four groups (small + sorted, small + unsorted, large + sorted, large + unsorted) are shown in Figure 43, which indicates a significant main effect of data separation on task completion time, and slight or insignificant effect of sorting. This is confirmed by the within-subject with blocking ANOVA analysis (Table 6) which indicates a significant effect of data separation ($p = 0.0001$), but insignificant effect in sorting ($p = 0.49$) without any evidence of factor interaction ($p = 0.26$).

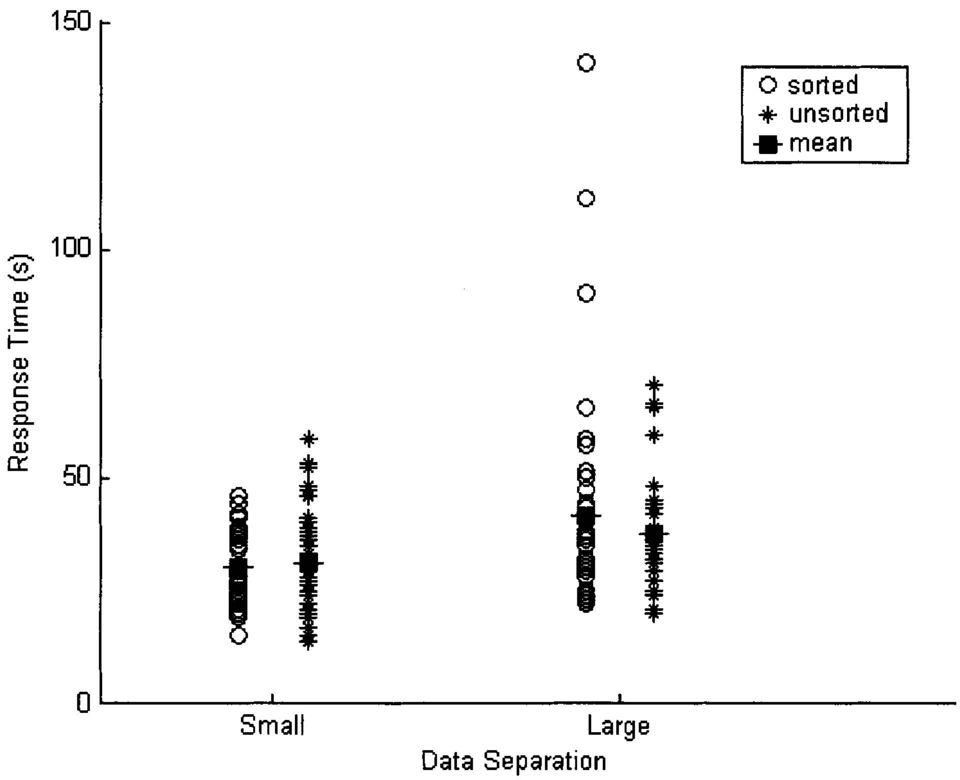


Figure 43. Scatter diagram for the linear display time results.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Sorting	80.6	1	80.6	0.47	0.49
Separation	3640.3	1	3640.3	21.4	<0.0001
Sorting*Separation	219.0	1	219.0	1.29	0.26
Participant (Block)	8227.5	10	822.8		
Error	27495.5	162	169.7		
Total	39717	175			

Table 6. ANOVA table for the linear display time results.

The statistics of the linear results are summarized in Table 7.

	Small + Sorted	**Small + Unsorted**	**Large + Sorted**	**Large + Unsorted**
Median	28.5	28.	35.5	36.
Mean	29.9	30.9	41.3	37.7
Std Deviation	8.3	9.8	23.2	11.6
95% CI	(27.4, 32.4)	(28.0, 33.7)	(34.3, 48.4)	(33.9, 41.4)

Table 7. Statistic for the linear time results.

All results are for task completion time in seconds.

2. Accuracy Results

The effects of data separation and sorting do not seem to be significant on task accuracy from the Freidman results (Table 8). Since the accuracy for each individual trial is either 0 (incorrect), or 1 (correct), scatter diagrams are not well-suited to display the data. Instead, a plot of means with error bars of one standard deviation below and above the mean is used (Figure 44). A summary of the statistics is listed in Table 9.

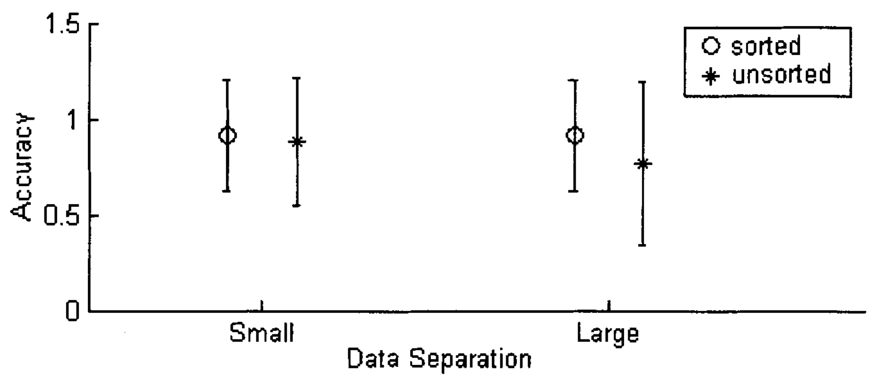


Figure 44. Plot for accuracy measurement in linear display results.

The means of the data sets are indicated by the two symbols, while the error are depicted by the error bars. The error bars are a distance of one standard deviation above and below the mean.

Source	Sum Sq.	d.f.	Mean Sq.	χ^2	Prob> χ^2
Sorting	111.4	1	111.4	3.3	0.070
Sorting*Separation	17.8	1	17.8		
Error	1293.8	40	32.3		
Total	1423	43			

Source	Sum Sq.	d.f.	Mean Sq.	χ^2	Prob> χ^2
Separation	44	1	44	1.37	0.24
Sorting*Separation	44	1	44		
Error	1257	40	31.4		
Total	1345	43			

Table 8. Freidman test for the accuracy measurement for the linear displays.

	Small + Sorted	Small + Unsorted	Large + Sorted	Large + Unsorted
Median	1.	1.	1.	1.
Mean	0.91	0.88	0.91	0.77
Std Deviation	0.29	0.33	0.29	0.43
95% CI	(0.82, 1.00)	(0.78, 0.97)	(0.82, 1.00)	(0.63, 0.91)

Table 9. Task accuracy statistics for the linear display results.

Experiment 2:

The Effect of Category Sorting using the Hierarchical Presentation Technique

1. Time Results

A scatter diagram of the results is shown in Figure 45, and it can be seen that there is no evidence to suggest an effect of category sorting of the data points requested by the logic rules. This is confirmed by the results of the ANOVA (Table 10, $p = 0.73$).

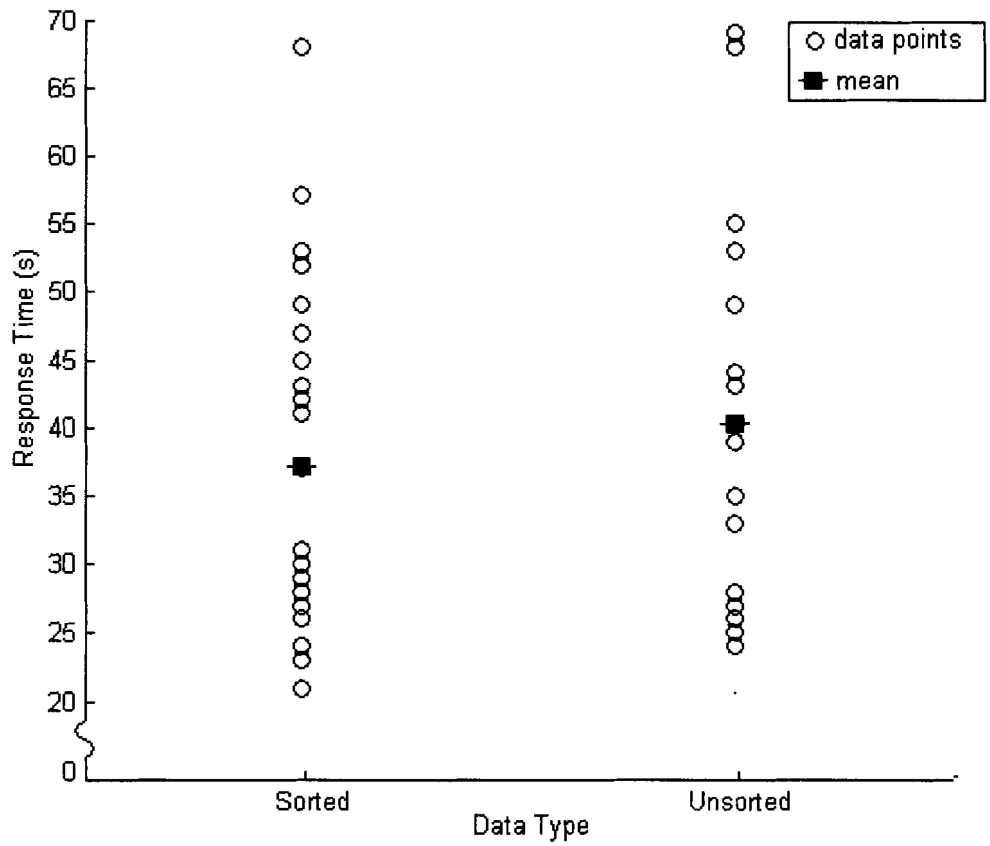


Figure 45. Scatter diagram of sorted and unsorted hierarchical time results.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Sorting	13.7	1	13.7	0.12	0.73
Participant (Block)	3940.5	10	394.1		
Error	3575.6	32	111.7		
Total	7516.2	43			

Table 10. ANOVA result for hierarchical time results with category sorting as the single factor.

	Sorted	Unsorted
Median	34	37
Mean	37.1	40.2
Std Deviation	12.2	15.0
95% CI	(32.4, 41.8)	(32.2, 48.3)

Table 11. Time statistic for hierarchical display results for sorting.

Data points requested by the logic rules were taken from two adjacent categories in the hierarchical display. The effect of category sorting was studied. All results are for task completion time in seconds.

2. Accuracy Results

Unlike to the time result, the Kruskal-Wallis results (p = 0.03, Table 12) suggest a main effect of category sorting in the listing of the data required. A plot of means with error bars is shown in Figure 46 and a summary of the statistics is listed in Table 13.

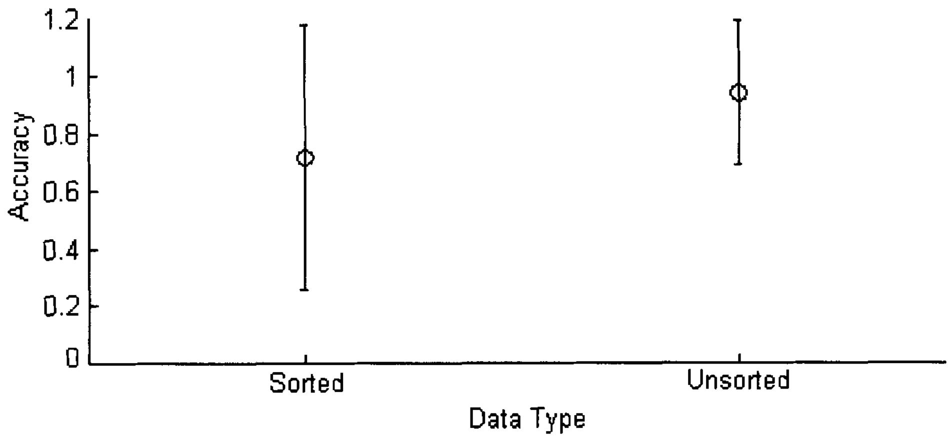


Figure 46. Plot for sorting effect in hierarchical displays.

The means of the data sets are indicated by the two symbols, while the error are depicted by the error bars. The error bars are a distance of one standard deviation above and below the mean. Each bar is symmetric two standard deviations long.

Source	Sum Sq.	d.f.	Mean Sq.	χ^2	Prob> χ^2
Sorting	127.7	1	127.7	4.47	0.034
Error	472.3	20	23.6		
Total	600	21			

Table 12. Kruskal-Wallis test for accuracy measurements in hierarchical displays with sorting as the single factor..

	Sorted	Unsorted
Median	1	1
Mean	0.71	0.94
Std Deviation	0.46	0.25
95% CI	(0.54, 0.89)	(0.80, 1.07)

Table 13. Task accuracy statistics for hierarchical display sorting effects.

Experiment 3:
The Effect of Data Category Distribution using the Hierarchical Presentation Technique

1. Time Results

Based on the results of Experiments 1 and 2, the effects of category sorting in groups 1 and 2 are found to be insignificant. When data requested is contained in a single category, the situation is analogous to that in a small-separation linear display. Based on results of Experiment 1 where the effect of sorting was found to be insignificant for small-separation linear displays, it can be postulated that the effect of sorting for single-category hierarchical displays would be insignificant. Similarly, the results of Experiment 2 indicate the effect of sorting was found to be insignificant in double-category distributed hierarchical displays. In other words, the results obtained in the five groups can be considered comparable despite the inconsistent sorting (i.e., a mix of sorted and unsorted data required by the tasks) in groups 1 and 2.

The scatter diagram for the five data category distribution group is shown in Figure 47.

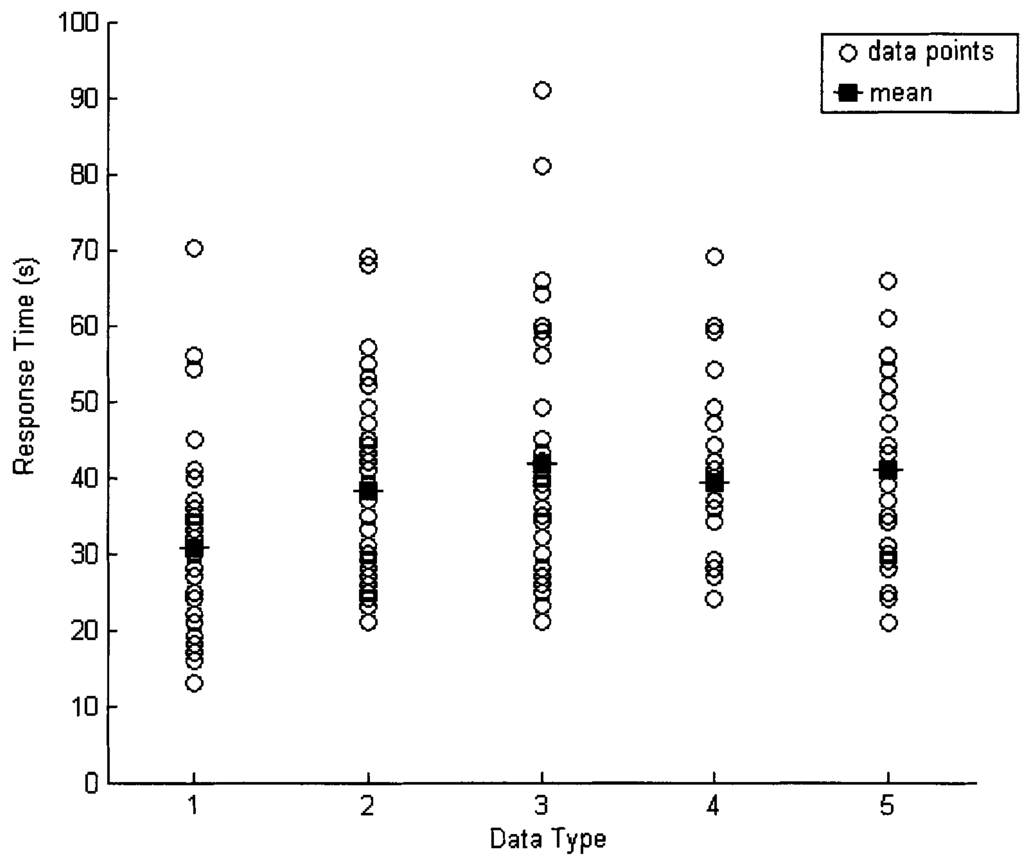


Figure 47. Scatter diagram of time results for the five data category distribution groups in the hierarchical display.

The scatter diagram above (Figure 47) shows that if data are confined within a single category, the performance time is significantly lower than if the data are more scattered, as corroborated by the ANOVA results ($p < 0.0001$, Table 14). A multiple comparison test+ showed that Group1 data is significantly different from the rest.

+ This was done using MatLab's `multcompare` function.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Separation	3313.0	4	828.3	7.77	<0.0001
Participant (Block)	12342.5	10	1234.3		
Error	17798.46	167	110.4		
Total	33361.8	181			

Table 14. ANOVA time results for the five data category distribution groups in the hierarchical display.

The statistics for the five groups are shown as follows:

Group	1	2	3	4	5
Mean	30.7	38.2	41.8	39.4	40.9
Std Deviation	11.0	13.2	15.2	12.0	13.3
Median	28	36	39	37	38
95% CI	(27.4, 34.0)	(34.2, 42.3)	(37.0, 46.6)	(34.8, 44.1)	(35.3, 46.5)

Table 15. Statistics of the time results for the five data category distribution groups in the hierarchical display. All results are for task completion time in seconds.

2. Accuracy Results

Since Experiment 2 discovered a main effect of sorting, the data in Group 2 of the hierarchical data cannot be combined. The task accuracy results for the six category-distribution groups are quite similar, except for the second unsorted, forth and fifth group, where the means are significantly higher (Figure 48, $p = 0.005$). The statistics for the accuracy measurements are summarized in Table 15.

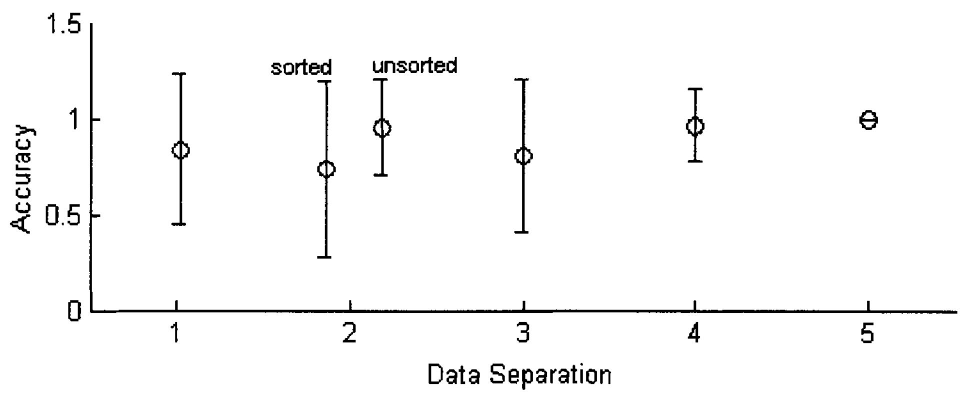


Figure 48. Scatter diagram for accuracy measurements in data category distribution in hierarchical display results.

Source	Sum Sq.	d.f.	Mean Sq.	χ^2	Prob> χ^2
Separation	4085.6	5	817.1	16.8	0.005
Error	11706.4	60	195.1		
Total	15792	65			

Table 16. Kruskal-Wallis test results for the five data category distribution groups in the hierarchical display.

Group	1		2	3	4	5
Median	1	1	1	1	1	1
Mean	0.82	0.71	0.94	0.81	0.96	1
Std Deviation	0.39	0.46	0.25	0.40	0.19	0
95% CI	(0.70, 0.94)	(0.54, 0.89)	(0.80, 1.07)	(0.69, 0.93)	(0.89, 1.04)	(1, 1)

Table 17. Task accuracy statistic for the data category distribution groups in hierarchical display results.

Questionnaire Results

There were two parts to the questionnaire: five rating questions using a 7-point numerical scale with 1 being strongly disagree and 7 being strongly agree, and four open-ended questions. For the rating questions, participants did not have a preference for the presentation techniques in performing the study task overall (median = 4). While it may seem that the participants favoured the linear display in locating the data points and in matching the options using the linear (both medians = 6) over the hierarchical displays (both medians = 5), the corresponding linear and hierarchical groups of results are not significantly different based on the Wilcoxon signed rank test[5]. This result is very encouraging considering that 75% of the hierarchical display rules required searching for data among two to five categories, and 63% of the rules were unsorted

---

[5]Wilcoxon signed rank test performs a paired, two-sided test of the hypothesis that the difference between the matched samples in the two groups comes from a distribution whose median is zero.

categorically. In the linear case, 50% of the data points required by the rules were contained within a single screen, and 50% were sorted alphabetically.

For the open-ended questions, the most *positive* aspects of the linear displays were found to be:

- No need to understand the meaning of the data name, or even read the whole name to locate the data point
- Less disruptive when the task requires a heavy cognitive load

while the most *negative* aspects were:

- Visual clutter given the small screen
- Excessive scrolling required
- Easy to confuse adjacent data names when they were similar

For the hierarchical displays, the most *positive* aspects were considered to be:

- Shorter list once the category was determined.
- Visually clear and compact with fewer distracting items
- Easier to remember when data fit into the same category

while the most *negative* ones were:

- The need to understand the data name to classify into a category
- Extra clicks required in opening and closing categories
- Items may be initially hidden

8.2.6 Discussion

The linear results show that there is a significant data separation effect on performance time ($p < 0.0001$), with a difference of approximately 6 seconds on a 36-second task. In other words, when the data requested could be seen on a single screen, it is on average 16% faster to perform the task. If the data could not fit onto a single screen, the performance cost when compared to the averaged time was found to be about 4% when the requested data were unsorted, or 14% when sorted. This apparently confusing result is caused by two outliers in the large + sorted group. In both cases, the outlier was caused by a single large time measurement (> 100 seconds). It is believed that the participants probably lost track of the answers for the logic statements, and started all over again. These two outliers inflate the mean and standard deviation of the large + sorted group. If those two sets of data were removed from the analysis, the two large groups would have similar means (about 37 seconds) and standard deviations (about 15 seconds), and the performance cost would have been about 8% based on a smaller averaged task time of 34.4 seconds. The conclusions on the effects of sorting and data separation would remain the same.

In short, sorting does not seem to affect performance time. In fact, similar conclusions can be drawn for the effect of data category sorting using the hierarchical presentation technique, where no evidence could be found on the effect of sorting.

Time results on category data distribution using hierarchical displays demonstrated a significant effect on task performance ($p < 0.0001$). When the data required for the task were contained in a single category, the performance benefit was found to be approximately 15%, similar to that found in small-separation linear displays. When the data required were distributed among two adjacent categories, the performance mean was slightly above average (5%), while with category distribution of three or more groups, there is a 12% performance cost, which is similar to that of the large-separated, sorted linear group. As in the case of the large-separated, sorted linear group, two outliers in the third group inflate the averaged performance time. However, taking these two outliers out of the analysis would not reduce the percentage performance cost since the averaged task time would also be reduced by a similar proportion. It is interesting to note that the performance cost was found to be similar in groups where data scattered among three or more categories. It suggests the cost penalty plateaus for data scattered among three or more group when using the hierarchical presentation technique.

The performance time results can therefore be summarized as follows. On the side of benefit, the hierarchical and linear displays were comparable in terms of maximum benefit, in situations when the data points requested were well suited to the presentation technique (i.e., when the data required for the task were contained in a single category for hierarchical displays, or when they fit on a single screen for linear displays). On the side of cost, despite the numerical difference in performance costs between large-separated linear display and large- separated hierarchical display, the difference is statistically insignificant based on ANOVA results, even before the outliers were removed ($p = 0.2466$). In other words, there is insufficient evidence to conclude a difference in maximal potential performance times between the two presentation techniques.

The accuracy results in Experiment 1 indicate that there is insufficient evidence for main effects of sorting and data separation on performance accuracy in the linear display. In the hierarchical displays (experiments 2 and 3), there was evidence to suggest main effects of sorting (Experiment 2, $p = 0.03$) and data category distribution (Experiment 3, $p = 0.005$). In Experiment 2, participants produced more accurate answers when the data requested by the rules were not sorted; while in Experiment 3, higher accuracy were obtained when the data requested in the rules were highly distributed among the categories (groups 4 and 5). In fact, for group 5 where every data item retrieved was from a different category, all the participants obtained perfect scores. When the large-separated hierarchical accuracy results were compared to the large-separated linear results, the numerical difference is found to be statistically significant ($p = 0.03$). These results may indicate that the participants were more careful in difficult tasks.

As indicated by the comments made by participants in the post-test questionnaire, the major costs in using the hierarchical displays are the needs to classify the data point of interest and to open and close the category to bring the data points in view. With tasks that already stressed the participant's memory, the extra cognitive load required in categorization was acutely felt, and the act of opening and closing of categories was felt to be disruptive. Nonetheless, despite these costs, the maximum performance cost of using hierarchical displays was found to be comparable to that in the linear display when the data points requested were unsorted with large separation.

In summary, the use of hierarchical presentation technique on a small display should be beneficial under the following circumstances:

- *When the majority of data required for the task is likely to belong to one or two categories.* Since the performance cost in the case of required data being scattered among three or more category was found to be comparable to large data separation in linear displays, hierarchical displays should provide better performance benefit than linear displays on average, except of course in cases where the majority of the data required for the task are scattered over three or more categories.

- *When the categorization scheme is simple, or well known to the user.* As commented by some participants where English was not their mother tongue, it required thinking and memory to classify the data points. Even for native English speakers, the memory stress in this task rendered the categorization at times conscious and deliberate.

- *When the list is long compared to the number of categories.* This was not tested directly in the study, but based on comments of participants in the post-test questionnaire, many felt that if the list were long, it would be much more difficult to search using the linear display. Similarly, if there were too many categories, it was postulated to be difficult to locate the data using the hierarchical display. This is very similar to the depth/breadth tradeoff in menu structures (Kiger, 1984; Landauer & Nachbar, 1985), and the horizontal depth study in mobile phones (Chae & Kim, 2003)

8.3 Conclusions on the Use of Hierarchical Presentation Technique

The maximal potential benefit and cost of using hierarchical displays were found to be comparable to those of linear display in terms of performance time. In other words, the best achievable performance time for the linear displays is comparable to that of the hierarchical ones, and the worst case for performance time for the linear displays is also comparable to the hierarchical ones. However, participants obtained higher accuracies when the data listed in the rules were either unsorted categorically, or highly distributed among the categories when using the hierarchical displays. If the majority of the data required for the task can be contained in one or two categories, it is very likely that the hierarchical display would better support the task in terms of performance time. In addition to data separation, other considerations for choosing among the two presentation techniques include the degree of familiarity of users to the categorization scheme, the length of the data list, and the number of categories. Based on participant comments, the hierarchical presentation technique would be favoured if the users were familiar with the organizational scheme, or when the scheme was very simple; or if the list were long; or if the number of categories involved were small. Further investigations are required to determine the optimal list length and category number for the two presentation techniques.

For the case of the medical laboratory result reporting system, the hierarchical presentation technique is well suited for the application. First, since most current laboratory reports are grouped hierarchically, physicians expect data grouping. Second, since the organizational scheme chosen is based on the ones provided by most laboratories, the physicians are likely to be extremely familiar with it. Third, the slight modifications of the organizational scheme made based on anatomical and physiological systems will likely help to concentrate interesting data to one or two categories. Forth, since laboratory tests are used mostly as tools to rule-in or rule-out hypotheses, it is likely that the tests ordered would fall under a small number of categories when compared to the number of tests ordered. For these reasons, the hierarchical presentation technique is an appropriate choice as the main method in presenting laboratory test results.

9 Future Work

Follow-up work based on the current thesis can be divided into two areas: tasks that develop the prototype into a commercial product, and tasks that extend the interesting results found in the comparative user study.

9.1 The Medical Laboratory Report Application

In addition to the design plan 3 outlined in §6.2.5, a number of issues have to be addressed before the application can be commercialized.

9.1.1 Data Sharing and Security.

What kind of accessibility is acceptable for security?

The accessibility/security tradeoffs have to be carefully considered. Initially, a number of platforms were investigated to 'web-enable' the application including Java (J2ME) and Microsoft's .NET. However, a list of obstacles prompted the rejection of this idea. First, existing handheld devices are too slow to accommodate the designed displays using available web-browsers. The time required for downloading the graphical interface elements and the laboratory report over the Internet rendered the application unusable. Second, initial responses from potential clients indicate a general reluctance to expose patient information over the Internet, even though the data can be encrypted. Third, the security requirements of an Internet broadcast would impose a severe workload on the IT department. Due to these problems, the application is expected to be restricted to the hospital's wireless intranet. However, the design has taken potential future upgrades to the Internet into account.

9.1.2 Interface with the Laboratory Information System.

How does the system integrate with the hospital's laboratory information system?

Since this is specific for each hospital, it will be implemented when a client has been identified. The current system has an interface exposed for future connection.

9.1.3 Audit Trail

An audit trail recording events like report received and opened is needed for monitoring and legal reasons.

9.1.4 Other Features

Based on feedbacks from physicians from the observational user study and the initial user profile survey, a number of features are considered desirable in a laboratory reporting system. These features are either not related to the problem of information visualization or are too extensive to be covered by the current project.

1. ***Electronic Notification.*** Since physicians may not be able to check laboratory results as soon as they are available, it is important to alert the physicians of severely abnormal tests that require immediate attention. At the moment, the laboratory may call the nursing station to which the patient belongs, or hospitals may page physicians. The former solution is found to be problematic. Messages are hand-written on pieces of paper, or on the patient's chart, which may contain transcription error, handwriting ambiguities, or may not even reach the attending physician in a timely manner. Electronic paging is a more direct and error-proof approach.

2. ***Sharing of Reports and Views.*** When conferring with other medical professionals, it may be necessary to share laboratory information, either in the form of a current status snapshot, or as chronological disease profiles. Being able to transport these reports, either electronically or by paper, is therefore important in current health care practices.

3. ***Development and Incorporation into an Electronic Patient Record.*** The laboratory report is a small step towards the realization of the electronic patient record. Other supporting information includes patient's medical history, family history, physician's progress report and treatment summary.

9.2 The Depth/Breadth Tradeoffs in Hierarchical Displays on Small Screens

An interesting question sprung out of the comparative user study: What are the optimal depth/breath tradeoffs when using hierarchical displays on small screens? A similar question was studied in menu structures and menu-driver user interfaces. According to Paap and Roske-Hofstrand (1988, p. 205-235),

> The basic trade-off in hierarchical menu design is between depth and breadth... a flat, broad menu does not require the user to go through as many levels as a deep menu, thereby reducing the need for user navigation. At the same time, each node in the menu hierarchy becomes more complex in a flat menu structure, making the user choose between more options at each level. Since both navigation and decisions take time, neither too deep nor too broad trees are desirable in general.

Which begs the question of "how deep is deep, and how broad is broad?". Laudauer & Nachbar, (1985) believed that in general, breadth is favoured over depth; while others (Kiger, 1984; Miller, 1981) suggested intermediate branching factors (4 to 8) are better than the deepest or the broadest structures. According to Kiger (1984), the capacity of short-term memory may play a role in the limitation of breadth:

> Interestingly, the best tree structure used a menu breadth that falls within G.A. Miller's (1956) "seven plus or minus two" estimation of short-term memory capacity. This may indicate that optimal menu interfaces take advantage of this limitation by pushing the upper limit of this capacity on individual menu, and thereby reducing the depth requirements of the tree structure.

Paap & Cooke (1997) postulated that the limitation is due to crowding, where too many options are available to the users at the same time. The crowding effect is particularly pronounced when human processing time in selecting among the available options is much slower than the time needed to traverse the tree structures.

In the case of small displays, Chae and Kim (2003) studied horizontal depth, which is defined as the number of pages traversed before reaching the target (the right hand figure in Figure 49). This differs from vertical depth, which is defined as the number of screens scrolled within the same page before reaching the target (the left hand figure in Figure 49). They found that the optimal horizontal depth level is four. However, their study did not investigate vertical depth and breadth. It would therefore be interesting to investigate the breadth/depth tradeoffs in small displays.

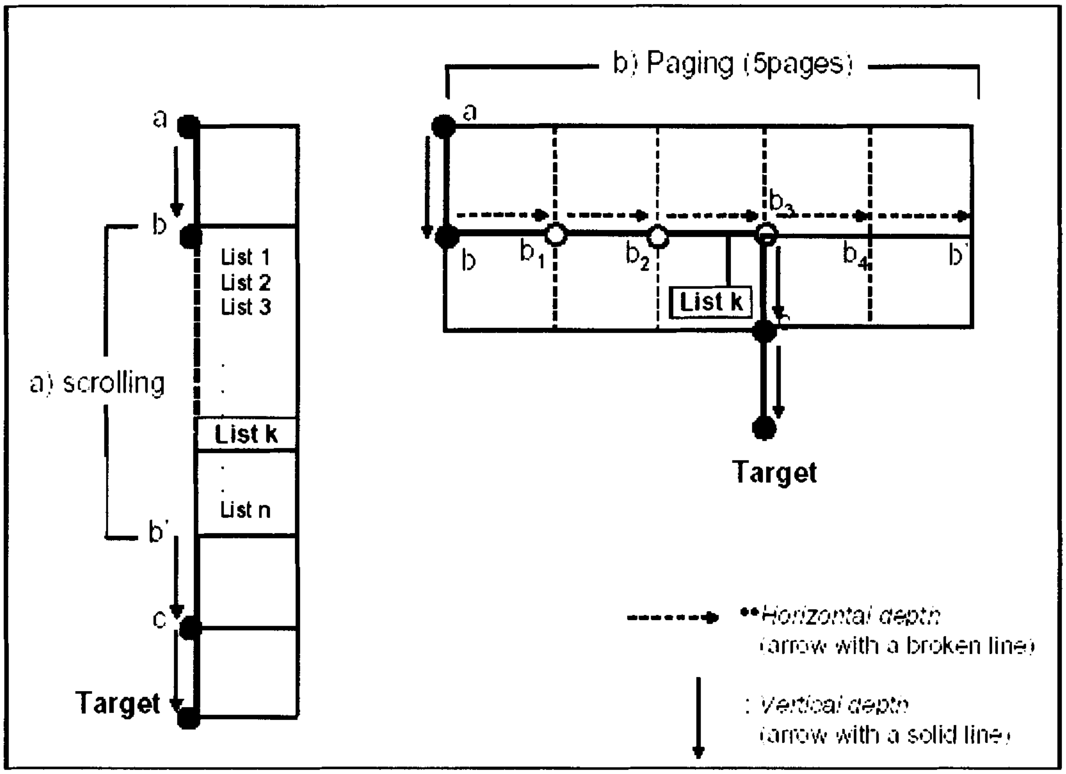


Figure 49. Horizontal and Vertical Depth.

Horizontal depth is defined as the number of pages traversed before reaching the target, while vertical depth is defined as the number of screens traversed before reaching the target. Copyright Chae & Kim, 2003, by permission.

10 Conclusions

This thesis proposed a design for medical laboratory results displayed on PDA-sized screens based on studies in the characteristics of the target users, the laboratory interpretation task, the characteristics of the PDAs, and the nature of the laboratory data. Findings from these areas provided design guidelines and goals for the prototype, with the main features being an anatomical overview for rapid patient status assessment, the use of hierarchical + elision presentation technique for better diagnostic support, and the provision of disease profiles for long-term patient management. A number of visualization techniques including perceptual layering, rapid vision, hierarchical + elision and "detail-on-demand" were used to reduce the severe design constraints of small devices.

The initial prototype was evaluated by eight physicians in an observational user study, leading to a number of design changes including decluttering of the anatomical overview display and tighter adherence to the grouping and linear ordering of laboratory tests on existing laboratory reports. A follow-up visit with four of the eight physicians confirmed the effectiveness of the changes, and provided further design ideas including the provision of disease-profiles for better patient management support. Further development in areas of security and interfacing to the laboratory's information system are required in commercializing the prototype.

The hierarchical + elision presentation technique used in the design to display a list of laboratory results was further studied to quantify its potential costs and benefits as a presentation technique for lists on small screen when compared to the more conventional linear technique. Despite the benefits of using the hierarchical presentation technique on larger desktop displays, the situation on smaller screen may be different. In fact, the comparative user study performed described in Chapter 8 indicates that on PDA-sized screens, the hierarchical + elision technique provided the same potential benefits as the conventional linear technique. On the side of costs, the maximum cost incurred by using the hierarchical + elision technique was also found to be comparable to the maximum cost incurred by the linear display. It is therefore important for the interface designer to understand the nature of the data displayed, the familiarity of the user to the organizational scheme and the user task when selecting an appropriate presentation technique to display lists of data on small screens.

Appendix A: Email Survey Questionnaire

1. What is the medium of delivery of your current laboratory report (paper, electronic)?
2. In general, what kind of tests do you find on a typical report?
3. What is/are useful in interpreting those test results? Are they different for different types of tests? I understand some tests do not have a numerical value, like microbiology, biopsy and radiology. What is important to you from those tests?
4. How many tests do you generally get per report?
5. How would you like to arrange those tests on the report? Alphabetically? By systems (cardiology, renal etc)? By severity of results (for example, the more ominous results are listed first), or other methods?
6. What do you like about the current reports?
7. What do you dislike about the current reports?
8. What do you wish to be included in the reports?
9. What kind of benefits do you expect from an online results reporting system? In other words, what should the system deliver before you will use it?

Appendix B: Observation User Study

Ethics Approval

SIMON FRASER UNIVERSITY

OFFICE OF RESEARCH ETHICS

BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6
Telephone: [illegible]
FAX: 604-268-[illegible]

May 31, 2003

Ms. Heidi Lam
Graduate Student
School of Engineering Science
Simon Fraser University

Dear Ms Lam:

Re: Subjective evaluation of a handheld medical laboratory result reporting system

I am pleased to inform you that the above referenced Request for Ethical Approval of Research has been approved on behalf of the Research Ethics Board. This approval is in effect for twenty-four months from the above date. Any changes in the procedures affecting interaction with human subjects should be reported to the Research Ethics Board. Significant changes will require the submission of a revised Request for Ethical Approval of Research. This approval is in effect only while you are a registered SFU student.

Your application has been categorized as "minimal risk" and approved by the Director, Office of Research Ethics, on behalf of the Research Ethics Board in accordance with University policy R20.01, http://www.sfu.ca/policies/research/r20-01.htm. The Board reviews and may amend decisions or subsequent amendments made independently by the Director, Chair or Deputy Chair at its regular monthly meetings.

"Minimal risk" occurs when potential subjects can reasonably be expected to regard the probability and magnitude of possible harms incurred by participating in the research to be no greater than those encountered by the subject in those aspects of his or her everyday life that relate to the research.

.../2

Page 2

Please note that it is the responsibility of the researcher, or the responsibility of the Student Supervisor if the researcher is a graduate student or undergraduate student, to maintain written or other forms of documented consent for a period of 1 year after the research has been completed.

Best wishes for success in this research.

Sincerely,

Dr. Hal Weinberg, Director
Office of Research Ethics

c: Dr. John Dill, Supervisor

/jmy

Questionnaire

This is to solicit subject user evaluation and reaction to the medical laboratory result reporting application run on a personal digital assistant.

			1	2	3	4	5	6	7		NA
1	It is easy to find the laboratory test result I need using this system.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
2	It is easy to understand the meaning of the results displayed using this system.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
3	It was easy to learn to use this system.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
4	The organization of information on the screens is clear.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
5	The information complements the way I think about the case presented in the study.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
6	This system has all the functions and capabilities I expect it to have.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
7	Overall, I believe this system will be more helpful to me in my patient management than the paper reports.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
8	Overall, I like using this system.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
			1	2	3	4	5	6	7		NA

List the three most **negative** aspect(s) of the system:

1. ______________________________
2. ______________________________
3. ______________________________

List the three most **positive** aspect(s) of the system:

1. ______________________________
2. ______________________________
3. ______________________________

Clinical Case Used in the Study

HISTORY

CHIEF COMPLAINT

The patient, Bob, relates an unexplained, sudden increase in thirst and appetite accompanied by abnormally frequent urination and weight loss of 3 pounds, starting approximately one month ago.

HISTORY OF PRESENT ILLNESS

Bob is a 45 y/o, moderately obese white male who presents with a one month history of unexplained, sudden onset of polydipsia, polyphagia, increased urinary frequency, and moderate weight loss. He denies pain on urination, blood in urine, change in stream, renal infection, or any recent changes in physical activity.

PAST MEDICAL HISTORY

Bob had an inguinal hernia repair at age 37. Medications include variable use of OTC pain medications, including aspirin and Ibuprofen, as well as occasional multivitamins. No known allergies, hypertension, pulmonary, thyroid, or kidney disease. Last clinic visit more than two years ago for an employment physical which was unremarkable except for being several pounds over his ideal weight.

REVIEW OF SYSTEMS

GENERAL

Bob relates a *weight loss* of about 3 pounds in the past month without dieting. In fact, he says that he has been eating more than usual. Bob also complains of feeling *fatigued* at the end of the day for about 3 months now, even though his job hasn't changed in years. His **appetite** has changed such that he's always hungry. He denies any weakness, chills, fever, or night sweats.

SKIN

Bob denies any color changes in his skin, problems with itching, bruising easily, red spots on his skin, changes in his moles, any infections, rashes, problems with hair or nails, or any cuts or scrapes.

HEMATOPOIETIC SYSTEM

Bob denies ever having anemia, has never received a transfusion, has never had spontaneous bleeding or excessive bleeding after tooth extraction, tonsillectomy, or other minor injury, has never noticed any enlarged or tender nodes or lumps, and has never craved paint or dirt.

HEAD & FACE

Bob denies a history of migraines or other headaches, and has never had any kind of facial pain or head/face trauma.

EARS

Bob denies any ear pain or discomfort, pus or discharge from his ears, ringing in his ears, or any loss of hearing.

EYES

Bob says that his **vision** hasn't been so good lately. His eyes started getting worse about 2 years ago. He notes some **blurring of vision**; when he drives, he notices things in the distance are blurry. Bob denies any eye pain, inflammation in or around his eyes, eye infections, trouble with seeing double, a blind spot, cataracts, or glaucoma. He doesn't wear glasses or contacts.

NOSE & SINUSES

Bob suffers from **post-nasal drip** every May when certain trees bloom. He denies nose bleeds, obstruction of his nose or sinuses, a discharge from his nose or sinuses, or any sinus pain.

MOUTH

Bob says that his **gums bleed** when he flosses and that he had his wisdom *teeth pulled* by a dentist when he was 33. He denies sores or cuts in his throat or mouth, any dental abscesses, a sore throat, hoarseness, or a sore tongue, and does not wear dentures.

PULMONARY SYSTEM

Bob has a slight *cough* upon waking but it goes away after a few minutes. He hasn't noticed a **change** in his **cough** for at least a year. Bob **smokes** about ½ pack a day and has done so for the past *10 years*. His last **chest x-ray** was about 2 years ago. His cough is not productive, and he has no trouble with wheezing, has never coughed up blood, denies chest pain when he coughs, does not have night sweats, has no occupational exposure to dusts or fumes, has never been exposed to anyone with TB, and does not frequently suffer from respiratory infections.

CARDIOVASCULAR SYSTEM

Bob denies any chest pain when he exercises or works hard, trouble breathing when he plays or works hard, trouble breathing at night, a need to use several pillows to get to sleep, fluid in his legs, palpitations, high blood pressure, having a murmur, varicose veins, any heart disease, rheumatic fever, calf pain after climbing stairs or exercising, or having ever fainted.

GASTROINTESTINAL SYSTEM

Bob's **appetite has changed** in that he's always hungry. He also reports a **change in his thirst**; he's always thirsty. He rarely has **heartburn**, usually when he overeats spicy food. His use of **alcoholic beverages** consists of a couple of beers a day after work. He had a **hernia** repaired when he was 33. Bob denies problems with nausea or vomiting, ever having vomited blood, any special dietary restrictions, trouble with gas, sour belching, trouble swallowing, or abdominal pain. He denies a history of jaundice, hemorrhoids, abnormal bowel movements, any change in bowel movement frequency, diarrhea, constipation, red blood in stool, tarry stools, change in bowel habits, use of laxatives, analgesic therapy, steroid therapy, or exposure to toxic substances.

URINARY TRACT

Bob noticed an **increase in the volume of urine** he produces; he's constantly in the bathroom. He also has to *urinate more at night*, getting up at least once in the middle of the night to urinate, whereas, before about a month ago, he never got up at night. He denies painful urination, unusual color in his urine, suddenly having to urinate, blood in his urine, kidney stones, trouble with his urinary stream, incontinence, enuresis, retention of urine, or trouble with bloating.

GENITAL TRACT

Bob denies any penile discharge or sore, testing positive for any sexually transmitted disease, any testicular pain or swelling, or a mass or anything abnormal in his scrotum.

MUSCULOSKELETAL SYSTEM

Bob has had right shoulder *joint stiffness* ever since he hurt it lifting some firewood out of his truck five years ago. He has **no limitation of motion** in his joints as long as he warms up first. He denies pain in his arms or legs, back pain, neck pain, joint swelling, joints feeling unusually warm, joint redness, joints making noise when he moves them, sprains, or deformities.

NERVOUS SYSTEM

Bob denies any convulsions, ever passing out, dizziness, vertigo, tremors, trouble walking, difficulty with speech, muscle atrophy, muscle tenderness, limp muscles, muscle weakness, muscle paralysis, any part of his body suddenly becoming numb, or a sudden sensation of burning, pricking, tickling, or tingling.

ENDOCRINE SYSTEM

Bob says that his **appetite suddenly increased**, in that, for the past month, he's hungry all of the time. His *thirst has also suddenly increased*; he's constantly thirsty. He denies having a goiter, tremors, heat or cold intolerance, abnormal sweating, changes in his voice, noticing that his eyes protrude more than normal, a sudden change in glove or shoe size, a change in his hair distribution, or a sudden change in his body contour.

PSYCHOLOGICAL STATUS

Bob denies any nervousness, irritability, memory loss, problems with depression, unusual fears, trouble getting to sleep, nightmares, impotence, or sexual disturbances. He doesn't oversleep on a regular basis and denies ever being incarcerated or arrested for violent behavior.

PHYSICAL EXAM

GENERAL

Well developed, moderately obese white male in no acute distress

Oral Temperature	37 °C
Rectal Temperature	NA
Respiratory Rate	17
Respiratory Rhythm	Regular
Pulse Rate	68
Pulse Rhythm	Regular
Height	5'10"
Weight	187 LB
Blood Pressure	140/90

SKIN

Hair normal with age-appropriate thinning. Nails normal with pink nail beds. Supple skin with minor sun damage. No lesions, inflammation, or swelling. Normal skin coloration with slight tan. No stasis. No ulceration. No excoriation. No edema. Normal skin turgor. Ambulatory ability normal.

HEAD

Skull normocephalic. No scalp lesions. No signs of head trauma. No bruits heard.

EARS

Hearing apparently normal; audio testing not done. Pinnae normal with no lesions. Canals patent. Tympanic membranes intact.

EYES

EOM normal. No ptosis or eyelid lesions. Uncorrected visual acuity 20/30 bilaterally. No xanthelasma. Conjunctivae clear w/o lesions. Sclerae intact. Cornea clear with no abrasions. Pupils PERRLA. No lens opacities. The fundal exam reveals changes indicative of moderate diabetic retinopathy. Note the light patches on the retinae.

NOSE

Nose intact, no external lesions. Nasal mucosa intact, no lesions. Slight septal deviation to the right. Turbinates visualized.

THROAT & MOUTH

Lip mucosa pink without lesions. Teeth moderately stained; several fillings noted. Gums pink to whitish near gumline with pronounced recession. Oral mucosa pink with no lesions. Tongue surface normal, muscle control normal. Tonsils present without redness or inflammation. Pharynx normal. Speech normal. Salivary glands normal. Trachea midline. Thyroid size normal.

CARDIOVASCULAR

No jugular venous pulsation. Carotid artery pulse normal. Apical impulse of normal location, size, force, and duration. No abnormal precordial movements. Cardiac auscultation reveals normal rate and rhythm. No murmurs, rubs, or clicks.

PULMONARY

Chest normal to inspection with no obvious deformity. Palpation normal with no masses, tenderness, or abnormalities noted. Lungs are normal to auscultation; no crackles, rubs, or other abnormal sounds. Lungs normal to percussion with no signs of consolidation or abnormal fluid levels.

ABDOMEN

Abdominal wall intact without distention, scars, or ventral hernia. Normal, uncircumcised penis without lesions. Scrotum intact without tears or lesions. Testes descended, normal in size. Palpation reveals no pain or tenderness, no evidence of ascites, no hepatosplenomegaly, no splenic tenderness, no rebound tenderness, and no masses. Auscultation normal; bowel sounds are present and not hyperactive.

RECTAL

Normal sphincter tone, no masses, no hemorrhoids, no occult blood. Heme test negative.

MUSCULOSKELETAL

No muscular weakness, tenderness, or atrophy. No peripheral edema. No spinal tenderness, abnormal curvature, or obvious deformity. Spinal mobility within normal limits. No joint deformities or masses.

NEUROLOGIC & PSYCHIATRIC

The patient appears alert, calm, appropriately dressed, talkative, and oriented x3. Normal affect with no signs of anxiety or depression. Motor activity level within normal limits. Patient is left handed. CN II – XII within normal limits. Neurological exam of extremities reveals normal motor function with strength and tone appropriate for patient's sex and age. No atrophy or fasciculations. Sensory exam reveals normal response to touch, pain, vibration, and position sense. Cerebellar exam reveals normal tandem walking, heel-knee, and finger-nose coordination. No tremor, chorea, or posturing. DTRs +2 and symmetrical all extremities. Negative Babinski sign.

STUDIES

Bob's rhythm strip reveals normal rate and rhythm. There is no evidence of past or current cardiac disease.

Bob's chest film is normal. No infiltrates and no hilar adenopathy.

LABORATORY

NOTE: This will be presented by the software

	Results	**Normal Values**
WBC	8,800	3.2-9.8X10^9/L
HCT	48	39-49%
HGB	14	12-15 g/dl
Platelets	250	130-400K/ml
HBA_{1C}	9.0%	4.3-5.8%
Sodium	140	135-147 mmol/L
Potassium	4.2	3.5-5.0 mmol/L
Chloride	100	95-105 mmol/L
BUN	32	8-18 mg/dl
Creatinine	1.2	0.6-1.2 mg/dl
Glucose	330	70-110 mg/dl
CO2	28	24-34 mEq/L
Prothrombin TIme (PT)	12	10-12 secs
Partial Thromboplastin Time (PTT)	30	25-41 secs
Cholesterol	160	0-200 mg/dl
LDL	120	40-70 mg/dl
HDL	40	50-190 mg/dl

UA	
Glucose	Negative
Protein	Trace
Ketones	Trace
Specific Gravity	Normal
Microscopic	Trace

ABG	
pH	7.4
pO_2	88
pCO_2	40
Bicarb	19
O_2 Sat	95%

DIAGNOSIS

Bob is suffering from type 2 diabetes or Diabetes Mellitus. However, it's possible that a patient with a similar presentation has impaired glucose tolerance, or is a secondary diabetic with hemochromatosis, Cushing's Syndrome, acromegaly, pheochromocytoma, glucagonoma, or other endocrine disorder.

Remember that it's also possible that laboratory findings can be in error, or there can be drug effects, inadequate diet, stress, anxiety, trauma, or concomitant illnesses that can create spurious laboratory results.

Appendix C: Comparative User Study Participant Materials

Ethics Approval

SIMON FRASER UNIVERSITY

OFFICE OF RESEARCH ETHICS

BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6
Telephone: 604-291-3447
FAX: 604-268-6785

September 2, 2003

Ms. Heidi Lam
Graduate Student
School of Engineering Science
Simon Fraser University

Dear Ms. Lam:

Re: Comparative user study of a handheld medical laboratory result reporting system

I am pleased to inform you that the above referenced Request for Ethical Approval of Research has been approved on behalf of the Research Ethics Board. This approval is in effect for twenty-four months from the above date. Any changes in the procedures affecting interaction with human subjects should be reported to the Research Ethics Board. Significant changes will require the submission of a revised Request for Ethical Approval of Research. This approval is in effect only while you are a registered SFU student.

Your application has been categorized as 'minimal risk" and approved by the Director, Office of Research Ethics, on behalf of the Research Ethics Board in accordance with University policy R20.0, http://www.sfu.ca/policies/research/r20-01.htm. The Board reviews and may amend decisions or subsequent amendments made independently by the Director, Chair or Deputy Chair at its regular monthly meetings.

"Minimal risk" occurs when potential subjects can reasonably be expected to regard the probability and magnitude of possible harms incurred by participating in the research to be no greater than those encountered by the subject in those aspects of his or her everyday life that relate to the research.

.../2

Page 2

Please note that it is the responsibility of the researcher, or the responsibility of the Student Supervisor if the researcher is a graduate student or undergraduate student, to maintain written or other forms of documented consent for a period of 1 year after the research has been completed.

Best wishes for success in this research.

Sincerely,

Dr. Hal Weinberg, Director
Office of Research Ethics

c: Dr. John Dill, Supervisor

/jmy

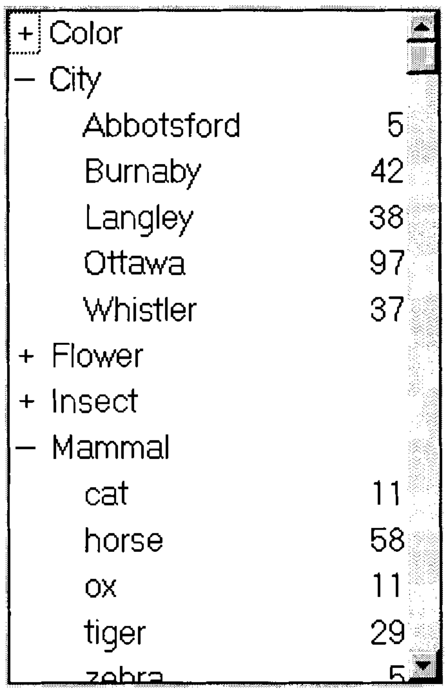
Instructions for the Study

Goal

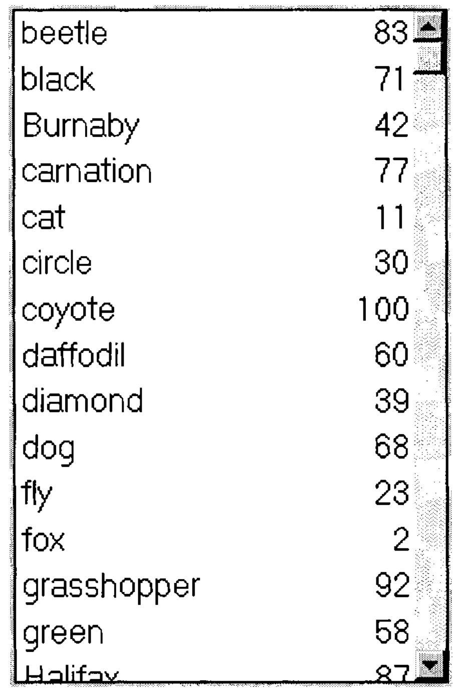
In this study, you will determine the logic state of a number of statements based on random data points that are displayed either hierarchically, or linearly in a small area on the computer screen.

Displays

Examples of the displays are as follows,



Hierarchical Display



Linear Display

In the *hierarchical* display, data points are grouped according to the category type. There are six types in this study: Color, City, Flower, Insect, Mammal and Shape:

Category	**Member Items**
Color	black, blue, green. magenta, orange, pink, purple, red, white, yellow
City	Abbotsford, Burnaby, Halifax, Kelowna, Langley, Ottawa, Surrey, Toronto, Vancouver, Whistler
Flower	buttercup, carnation, chrysanthemum, daffodil, iris, orchid, primrose, rose, sunflower, tulip
Insect	ant, beetle, cockroach, dragonfly, grasshopper, fly, mosquito, snail, spider, whitefly
Mammal	cat, coyote, dog, fox, horse, ox, pig, sheep, tiger, zebra
Shape	circle, diamond, hexagon, pentagon, rectangle, rhombus, octagon, parallelogram, square, triangle

Data points are arranged alphabetically within each group. Groups can be expanded or collapsed by clicking the "+" /"—" sign, or the group title.

For the *linear* display, data points are arranged alphabetically using the entire data name.

For both displays, scroll bars are used for scrolling up and down the lists.

Task

Each data point consists of a name (e.g., cat) with an associated value (e.g., 11). An example of the logic statement is:

```
Cat > 12
```

Your need determine if the statement is `TRUE`, or `FALSE` for each data set displayed. To evaluate the statement, you need to find the value of data point specified among the data displayed.

Each task will have a five such statements based on a different data set. Your goal is to determine the logic states of all of the statements, and match it to one of the five choices. The choices are in the form of "T T F T F", and differ from task to task.

The study procedure is as follows:

1. *Practice session*, consisting of 5 questions.

2. *Actual study session*, consisting of 32 questions. There will be a 2-minute break after 16 questions. You can also take breaks in between questions. The clock will start after you press "Next" and end with your last selection of the answer.

3. *Questionnaire,* consisting of 9 questions to solicit your subjective impression of the presentation techniques.

Note

1. You are encouraged to be as accurate and as fast as possible. However, please be certain of the answer before submitting it. The amount of time required to complete each question, along with the answer submitted, will be recorded.

2. It is easiest if you determine and think-aloud the logic state of each statement sequentially before considering the answer options. It will help you remember the logic states of the statements when choosing an answer.

Questionnaire

			1	2	3	4	5	6	7		NA
1	It is easy to find the necessary data using the **linear** display.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
2	It is easy to find the necessary data using the **hierarchical** display.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
3	It is easy to match the statements using the **linear** display.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
4	It is easy to match the statements using the **hierarchical** display.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
5	I prefer using the hierarchical than the linear display in doing the study task.	strongly disagree	O	O	O	O	O	O	O	strongly agree	O
			1	2	3	4	5	6	7		NA

List the three most **negative** aspect(s) of the **linear** display when performing the study task:
1. ______
2. ______
3. ______

List the three most **positive** aspect(s) of the **linear** display when performing the study task:
1.
2. ______
3. ______

List ~~the three most **negative** aspect(s) of the **hierarchical** display when performing the study task~~:
1. ______
2. ______
3. ______

List the three most **positive** aspect(s) of the **hierarchical** display when performing the study task:

1. ____
2. ____
3. ____

References

Altshuler, C. H. (1994). Data utilization, not data acquisition, is the main problem. *Clinical Chemistry*, 40, 1616-20.

Arocha, J. F., Patel, V. L., Patel, Y. C. (1993). Hypothesis generation and the coordination of theory and evidence in medical diagnostic reasoning. *Medical Decision Making*, 13, 198-211.

Appleby, C. (1997). Web-o-matic isn't automatic—yet. Internet technology hasn't broken the barrier between doctors and computers. *Hospitals and Health Networks*, 71(22), 30-1, 34.

Bartram, L., Ho, A., Dill, J., Henigman, F. (1995). The continuous zoom: a constraint fisheye technique for viewing and navigation large information spaces. *Proceedings of UIST '95*, ACM, pp. 201-215.

Bauer, B., Jolicoeur, P., Cowan, W. B. (1996). Distracter heterogeneity versus linear separability in colour visual search. *Perception*, 25, 1281-1294.

Bandolier. (2004) The same as yesterday, laddie!: inappropriate laboratory tests. Retrieved February 24, 2004, from http://www.jr2.ox.ac.uk/bandolier/band55/b55-2.html

Bonini, P., Plebani, M., Ceriotti, F., and Rubboli, F. (2002). Errors in Laboratory Medicine. *Clinical Chemistry*, *48*, 691 – 698.

Brooks, R. W. (1999). Using an intranet for physician desktop data consolidation. *Proceedings of the 32nd Hawaii International Conference on System Sciences*, pp. 1-14.

Brewer, W. F. (1987). Schemas versus mental models in human memory. In P. Morris (Ed.), *Modeling cognition* (pp. 187-197). New York: Wiley.

Buyukkokten, O., Molina, H. G., Paepcke, A. (2001). Accordion summarization for end-game browsing on PDAs and cellular phones. In *CHI 2001*, pp. 213-220.

Cartmill, R. S. V., Thornton, J. G. (1992). Effect of presentation of partogram information on obstetric decision-making. *Lancet*, 339, 1520-22.

Chapanis, A. (1965). *Man-machine engineering*. London, United Kingdom: Tavistock Publications.

Chae, M., Kim, J. (2003). An empirical study on the breadth/depth tradeoffs in very small screens on mobile internet phones. *Proceedings of 2003 Americas Conference on Information Systems (AMCIS)*, In Press.

Chernoff, H. (1973). Using faces to represent points in k-dimensional space. *Journal of the American Statistical Association*, 68, 361-368.

Deshe, O. and Van Laar, D. L. (1999). In Brewster & Dunlop (Eds.), *Proceedings of the Second Workshop on Human Computer Interaction with Mobile Devices*. Retrieved February 2, 2003 from http://www.dcs.gla.ac.uk/~mark/research/workshops/mobile99

Dillion, A., Richardson., J., McKnight., C. (1990). The effect of display size and text splitting on reading lengthy text from the screen. *Behaviour and Information Technology*, 9(3), 215-217.

Duchinichy, R. L., Kwahk, J. (1983). Readability of text scrolled on visual display terminals as a function of window size. *Human Factors*, 25, 683-92.

eMarketer. (2001, March). *The eHealth Report*. Retrieved October 30, 2002, from http://www.the-infoshop.com/study/em8158_e_health_report.html

Ericcsson, K. A., Staszewski, J. J. (1989). Skilled memory and expertise: mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235-267). Hillsdale, HJ: Erlbaum.

Fisher, B., Dill, J. (2000). Application of theories of indexical cognition to a web-based workspace. In American Association for Artificial Intelligence (Ed.), *Smart Graphics: Papers from 2000 AAAI Spring Symposium*. London, New York: Springer.

Furnas, G. W. (1999). Effective View Navigation. In Card, Mackinlay, and Shneiderman(Eds), *Readings in Information Visualization: Using Vision to Think* (pp. 589-596). San Francisco, Califonia,: Morgan Kaufmann.

Garrard, Ch. S. (2000). Human-computer interactions: can computers improve the way doctors work? *Schweiz Med Wochenschr*, 130, 1557-63.

Goldschmidt, H. M. J., Lent, R. W. (1995). From Data to information: how to define the context? *Chemometrics and intelligent laboratory systems*; 28, 181-92.

Harris Interactive. (2001, August 15). *Computing in the Physician's Practice*. Retrieved October 30, 2003, from http://www.harrisinteractive.com/news/newsletters/healthnews/HI_HealthCareNews2001Vol1_iss25.pdf

Healey, C. G, Booth, K. S., Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Human Computer Interaction*, 3(2), 107-135.

Hornof, A. J. (2001). Visual Search and Mouse-Pointing in Labelled versus Unlabelled Two-Dimensional Visual Hierarchies, *TOCHI*, 8(3), 171-197.

Hooke, J. O. O., Gelsema, E. S., Wulkan, R. W., Leijnse, B. (1991). Graphical non-linear representation of multi-dimensional laboratory measurements in their clinical context. *Methods Information in Medicine*, 30, 138-44.

Hunt, E. (1989). Cognitive Science: Definition, status, and questions. *Annual Review of Psychology*, 40, 603-629.

Kamba, T., Elson, S., Harpold, T., Stamper, T., Piyawadee, N. (1996). Using small screen space more effectively. In *Proceedings of CHI 96*, pp. 383-390.

Kiger, J. I. (1984). The Depth/Breadth Tradeoff in the Design of Menu Driver User Interfaces. *International Journal of Man-Machine Studies*, 20, 201-213.

Koedinger, K. R., Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14, 511-550.

Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G. (1999). Improving web interaction on small displays. *In Proceedings of the WWW8 Conference*, pp. 51-59.

Lasky, M. S., McCracken, H., Bielski, V. (1998, September 9). Just how helpful are handhelds? Retrieved October 30, 2003, from http://www.cnn.com/TECH/computing/9809/09/palmtops.idg

Lamping J., R. Rao. P. Pirolli. (1995) A focus+context technique based on hyperbolic geometry for viewing large hierarchies. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 401-408.

Landauer, T. K. and Nachbar, D. W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth and width. In *Proceedings of the ACM CHI' 85*, pp. 73-78.

Leung Y., M. Apperley. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction*, 1(2), 126-160.

McNeil, B. J., Pauker, S. G., Sox, H. C., Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306, 1259-62.

Mayer, M., Wilkinson, I., Heikkinen, R., Ørntoft, T., Magid, E. (1998). Improved laboratory test selection and enhanced perception of test results as tools for cost-effective medicine. *Clinical Chemistry Laboratory Medicine*, 36(9), 683-690.

Mayer, M., Chou, D., Eytan, T. (2001). Unit-independent reporting of laboratory test results. *Clinical Chemistry Laboratory Medicine*, 39(1), 50-52.

Mayhew, D. (1999). *The usability engineering lifecycle : a practioner's handbook for user interface design*. San Francisco, USA: Morgan Kaufmann Publishers.

Miller, D. P. (1981). The depth/breadth tradeoff in hierarchical computer menus. In *Proceedings of the Human Factors Society*, pp. 296-300.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits of our capacity for information processing. *Psychology Review*, 63(2), 81-7.

Montgomery, D. C. (2001). *Design and analysis of experiments (5th ed.).* Hoboken, NJ: John Wiley & Sons, Inc.

Mullet, K. and Sano, D 1995. *designing visual interfaces: communication oriented techniques.* Prentice Hall, Englewood Cliffs, NJ.

Nielson, J. (1993). *Usability engineering.* Academic Press, London, United Kingdom.

Norman, G. R., Brooks, L. R., Allen, S. W. (1989). Recall by expert medical practitioners as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 1166-1174.

Nygren, E, Lind M., Johnson M., Sandblad B. (1992). The art of the obvious: automatically processed components of the task of reading frequently used documents. In *Proceedings of the ACM CHI' 92*, pp. 235-39.

Paap, K., and Cooke, N. (1997). Design of menus. In Helander, M. and Carliner, S (Eds). *Handbook of Human Computer Intercation.* Amsterdam, Netherlands: Elsevier Sciene B. V..

PalmSource. (2002). *Palm OS Programmer's Companion Volume I User Interface.* Retrieved June 10, 2003, from http://www.palmos.com/dev/support/docs/palmos/UserInterface.html

Parker, G., Franck, G., Ware, C. (1998). Visualizing of large nested graphs in 3D: navigation and interaction. *Journal of Visual Languages*, 9, 299-317.

Patel, V., Arocha, J., Kaufman, R. (1994). Diagnostic reasoning and medical experitise. *The Psychology of Learning and Motivation*, 31, 187-252.

Patel, V., Groen, G., Patel, Y. (1997). Cognitive aspects of clinical performance during patient workup: The role of medical expertise. *Advances in Health Sciences Education*, 2, 95-114.

Politser, P. (1986). How to make laboratory information more informative. *Clinical Chemistry*: 1510 -16.

Powsner, S. M., Tufte, E. R. (1994). Graphical summary of patient status. *Lancet*, 344, 386-389.

Rao, R., Card, S. K. (1994). The Table Lens: Merging graphical and symbolic representations in an interaction Focus + Context visualization for tabular information. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 318-322.

Rensink, R.A. (2002). Internal vs. external information in visual perception. In *Proceedings of the Second International Symposium on Smart Graphics*, pp. 63-70.

Roberston, G. C., Card, K. C., Mackinlay, J. D. (1989). The cognitive coprocessor architecture for interactive user interfaces. In *Proceedings of the 2nd annual ACM SIGGRAPH symposium on User interface software and technology*, pp. 10-18.

Robinson, A. H., Morrison, J.L., Muehrcke, P.C., Kimerlng, A.J., Guptill, S.C. (1995). *Elements of Cartography* (6th Ed.). New York: Wiley.

Schaffer, D., Z. Zuo, S. Greenberg, J. Dill, L. Bartram, S. Dubs, D. Roseman. (1996) Navigating hierarchically clustered networks through fisheye and full-zoom methods. *ACM TOCHI* 3(2), 162-188.

Staggers, N., Mills M. E. (1994). Nurse-computer interaction: staff performance outcomes. *Nursing Research*, 43, 144-50.

Shumway, J. M, A. Il Jacknowitz, M. A. Abate. (1990) Analysis of physician's, pharmacists', and nurses' attitudes toward the use of computers to access drug information. *Methods of Information in Medicine*, 29 (2), 99-103.

Schwartz, D. R., Howell, W. C. (1985). Optional stopping performance under graphical and numeric CRT formatting. *The Human Factor Society*, August, 433-444.

Tan, P. (1997). CZWeb: An aid to navigate the World-Wide Web. *Master (M.A.Sc.)thesis: Simon Fraser University*, www.sfu.ca

Triesman, A., Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, 95(1), 15-48.

University of Wisconsin Family Medicine. (2000). *Woman with high blood pressure. Diagnostics: What objective tests should now be done?* Retrived January 31, 2003, from http://www.fammed.wisc.edu/pcc/curr/hypertension/case3/labs.html

Van Laar, D. L. (2001). Psychological and cartographic principles for the production of visual layering effects in computer displays. *Displays*; 22, 125-135.

Verheij, R, Hooke, J. O. O., Bonke, B., Van Strik, R., Gelsema, E. S. (1997a). Evaluation of techniques for the presentation of laboratory data. I: Time needed for accurate interpretation. *Methods of Information in Medicine*, 36, 11-6.

Verheij, R, Hooke, J. O. O., Bonke, B., Van Strik, R., Gelsema, E. S. (1997b). Evaluation of techniques for the presentation of laboratory data. II: Accuracy of interpretation. *Methods of Information in Medicine*, 36, 17-9.

Ware, C. (2000). *Information visualization.* San Diego, USA: Academic Press.

Watters, C., Duffy, J., Duffy, K. (2002). *Moderating the effect of screen size.* Retrieved October 20, 2003 from http://www.cs.dal.ca/~watters/screensize/StudyFullPaper2.doc

Werner, M. (1995). A model for medical decision making and problem solving. *Clinical Chemistry*, 41(B), 1215-22.

Warde Medical Laboratory Test Catalog. (2002). Retrieved January 31, 2003, from http://www.wardelab.com/catalog.asp

Wickens, C. D. (1992). *Engineering psychology and human performance* (2d ed). New York: Harper Collins.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testings.* San Diego, CA: Academic Press.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53(3), 300-314.

Wood, M. (1968). Visual perception and map design. *Cartographic Journal*, 5, 54-64.

Wright, P., Threlfall, S. M. (1980). Readers' expectations about format influence the usability of an index. *Journal of Research in Communication Studies.* 2, 99-106.

Wright, P., Jansen, C., Wyatt, J. (1998). How to limit clinical errors in interpretation of data. *Lancet*, 352, 1539-1543.

Wyatt, J. C. (1994). Clinical data systems, part II: components and techniques. *Lancet*, 344, 1609-14.

Wyatt, J., Wright P. (1998). Design should help use of patients' data. *Lancet*, 352, 1375-78.

Young, D. W. (1984). What makes doctors use computers?. *Royal Society of Medicine*, 7, 663-7.