ICS REPORT 8401

SOFTWARE ENGINEERING
FOR USER INTERFACES

Stephen W. Draper and Donald A. Norman

INSTITUTE FOR COGNITIVE SCIENCE
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093

The research reported here was conducted under Contract N00014-79-C-0223, NR 667-437 with the Personnel and Training Research Programs of the Office of Naval Research, and was sponsored by the Office of Naval Research and a grant from: the System Development Foundation. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsoring agencies. Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ONR REPORT 8401
### Software Engineering for User Interfaces

Stephen W. Draper and Donald A. Norman

Center for Human Information Processing
Institute for Cognitive Science
University of California, San Diego
La Jolla, CA 92039

Personnel & Training (Code 442)
Office of Naval Research
800 N. Quincy St., Arlington, VA 22217

Approved for public release; distribution unlimited.

This research was also supported by a grant from the System Development Foundation.

**Key Words**
- software engineering
- human-machine interface
- design
- interface evaluation
- benchmarks
- debugging
- tradeoffs
- documentation
- user needs
- acceptance tests

**Abstract**

OVER
ABSTRACT

The discipline of Software Engineering can be extended in a natural way to deal with the issues raised by a systematic approach to the design of human-machine interfaces. Two main points are made: that the user should be treated as part of the system being designed, and that projects should be organized to take account of the current (small) state of a priori knowledge about how to design interfaces.

Because the principles of good user-interface design are not yet well specified (and not yet known), interfaces should be developed through an iterative process. This means that it is essential to develop tools for evaluation and debugging of the interface, much the same way as tools have been developed for the evaluation and debugging of program code. We need to develop methods of detecting bugs in the interface and of diagnosing their cause. The tools for testing interfaces should include measures of interface performance, acceptance tests, and benchmarks. Developing useful measures is a non-trivial task, but a start can and should be made.
SOFTWARE ENGINEERING FOR USER INTERFACES

Stephen W. Draper
and
Donald A. Norman

Institute for Cognitive Science C-015
University of California, San Diego
La Jolla, California 92039

ABSTRACT

The discipline of Software Engineering can be extended in a natural way to deal with the issues raised by a systematic approach to the design of human-machine interfaces. Two main points are made: that the user should be treated as part of the system being designed, and that projects should be organized to take account of the current (small) state of a priori knowledge about how to design interfaces.

Because the principles of good user-interface design are not yet well specified (and not yet known), interfaces should be developed through an iterative process. This means that it is essential to develop tools for evaluation and debugging of the interface, much the same way as tools have been developed for the evaluation and debugging of program code. We need to develop methods of detecting bugs in the interface and of diagnosing their cause. The tools for testing interfaces should include measures of interface performance, acceptance tests, and benchmarks. Developing useful measures is a non-trivial task, but a start can and should be made.

Introduction

The subject of this paper is the extension of Software Engineering to deal with the issues raised by the design of human-machine interfaces. To a large extent, all that is needed is to take the problem of engineering the user interface as seriously as any other part of software engineering and to apply to it the same kind of techniques, appropriately adapted. For instance, although the interface is implemented in software, it can be thought of as being "run" on human users. This means that we must modify our concept of a program bug to allow for part of the system to be a person; we must establish new performance criteria for the combined human-plus-interface system. Much of the thrust of this paper is simply to draw analogies with existing software practices in order to suggest how to support a progressive approach to interface design. We do not present detailed ideas on what interfaces should be like, but rather sketch some consequences for software engineering when interface design is taken seriously.

There are two themes in this paper: arguing by analogy in software engineering and extending to interface design; and arguing from the nature of the problems of interface design to requirements for an appropriate engineering discipline. The first part of the paper makes some general points, the second summarizes their consequences for the coding, documentation, debugging, and testing phases.

Software Engineering for Human-Computer Interface Design

Two Goals in Optimizing an Interface

We begin with a tradeoff that has clear parallels to aspects of software engineering, the tradeoff between two quite different broad aims for an interface:

- Achieving speed and convenience of use (power) for the practiced user;
- Achieving ease of learning and use (ELU).

These distinctions are related to the familiar novice-expert distinction. There is a clear analogy between these aims and the desire to optimize space and time in software performance. In general it is desirable to optimize both aims, but beyond a certain point, the engineer must choose a particular tradeoff between the two. This analogy holds quite closely. Not only is there a reasonable region where one aim trades off against the other, but, as the extremes, the programs can be generally unusable.

"We leave out of this paper any major discussion of what a human-computer interface ought to be, or what the general problems or principles are, because this topic is explored in numerous other papers by us and by the growing community of people who work in the field called Human-Computer Interaction. This field now has an annual meeting, several journals, and an ACM Special Interest Group (SIGCHI); we see no need to cover the material here. See, for example, the Proceedings of the 23rd ACM Conference on Human Factors in Computing Systems." This paper concentrates on the issues that can be applied to interface design from practices already common in Software Engineering.
Consider the extremes. Suppose we optimized program speed or interface power. Just as a fast program that requires more space than any customer can afford is in practice useless, so too is an interface that provides immensely fast and "powerful" interaction, but at the cost of requiring such a level of skill that no user will actually be able to use it because of the difficulty of learning it presents. Similarly, if we optimized program space or ELU, the program might be too slow to use regularly (no matter how economical of space) and the interface can be too laborious for a regular user to employ productively, no matter how safe, "friendly," helpful, and easy to learn. If we avoid the extremes of these dimensions, then there is a wide range of choices for the designer. The existence of programs of similar function on microprocessors and large mainframe computers is a strong reminder that we can tradeoff both in the space-time domain for the software and in the power/ELU domain for the user interface.

User and Program as Communicating Co-Routines

When a programmer implements a system with a user interface, he or she is not only defining what the machine will do but also defining what the user can do. The program and user can be thought of as co-routines, each communicating with one another. The programmer must explicitly or implicitly decide what actions and information will be available to the user. For any given interface, this point can be illustrated by drawing a flow chart of the user's part of the process. Although this point seems simple, it changes the fundamental (though usually unacknowledged) situation by which programming takes place. No longer are programmers engaged in the private task of getting the machine to do something for themselves—they are not even engaged in communicating a program to other programmers, a view intermittently voiced by purist programmers. Instead, the interface designer should be viewed as someone who must write a successful co-routine between user and machine, yet where the full specification of the user side of the co-routine is not well known (and may be variable); this is the real subject of the design, and this must become the conscious objective.

There are two consequences to this point, both discussed in greater detail later in the paper. First, we need languages that support this view, that is, that represent this co-routine directly. Second, the interaction needs testing and debugging on typical "hardware," hardware that includes representative users.

The State of User Interface Design

How much is known about how to design interfaces? We must answer this question in order to plan a sensible software engineering strategy. If enough is known to lay down detailed principles, then a top-down strategy might be followed; otherwise an iterative strategy must be adopted with emphasis on the testing and debugging phases.

Quantitative principles. Top-down principles of design should, in the ideal case, allow a design to be worked out in advance rather than entirely by trial and error. These tools need to be developed from a solid basis in experimental psychology, coupled with a good understanding of programming and software design. Not surprisingly, the number of groups capable of this work is limited and not much yet exists. There are now several initiatives whose goal is to provide quantitative principles for the design of human-computer interfaces. For example, one of us has shown how it is possible to develop a quantitative assessment of how design tradeoffs affect the User Satisfaction for an interface. Thus, in trading off time, information, and workspace, it is possible to compute the tradeoff space, showing the psychological impact of tradeoffs in these variables. A complementary approach is that of Card, Moran, and Newell who provide a set of quantitative tools for competing the operational parameters of interfaces.

The development of psychologically based, quantitative design tools is still in its infancy and much work remains to be done, both in development and in validation. If we can develop sufficient range and breadth of quantitative design tools, then we can look forward to a time when it will be possible to provide general principles and even tables of numerical values in design handbooks, allowing such design considerations as workspace size, display time, menu structure, command language design, pointing-verification, interface power, and ELU to be assessed at the design stage, without the need to build a test system. For the present, however, the designer, by and large, does not have quantitative data at hand.

Qualitative principles. In the absence of well-established quantitative design aids, we need methods that allow us to work with qualitative principles. A large number of qualitative principles exists, many in the form of "slogans," enshrining the designer to consider this factor or that, or to avoid this failing or that. These qualitative rules are often quite reasonable (which makes it especially discouraging to see how frequently even the most obvious and elementary principles are violated in existing systems), although if the design of a user interface is ever to proceed in a systematic fashion, we need to go beyond this level.

One major example of the attempt to base a system design on fundamental design principles, built in such a way as to capitalize on the user's existing knowledge is the design of the Karus Star system. This design attempted a systematic strategy based on principles of good human-computer interaction. Since then, of course, other systems have been developed along similar principles (in particular, the Apple Lisa system). The Star illustrates the delicate tradeoff decisions that must be faced: the attempt to optimize ELU led to degradation of performance, especially in speed of the system, as well as to a high selling-price.
Specifications for Interfaces

In defining and addressing a software project two questions must be tackled: what kind of specifications are reasonable for defining the project? What kind of divisions of the project into subtasks are likely to be viable? The same observation applies to both: specifications are determined at an early stage and must remain constant if massive re-writing is to be avoided.

Shel11 argues against the utility of a Structured Programming approach to user interface design. Shel1 points out that whatever the virtues of that approach, it depends on having a clear and fixed specification at the start. The technique becomes strained whenever the specifications change during the life of the software. Although in the ideal case a Structured Program practitioner would redo the whole design and implementation at every change in specification, in practice the investment in the existing code and design exerts an enormous pressure toward piecemeal change, i.e., toward iterative redesign. In areas where such shifts in specification are common, it seems better to face the fact that development will in practice be iterative. Shel1 suggests—and we agree—that interface design is one such area.

Several consequences follow from this point of view. First, it will be unwise to allow any details of a user interface to appear in the project specifications; for instance instead of giving a list of the commands to be implemented, specifications should rather be of the form discussed by Shneiderman10, who suggests instituting acceptance tests for interfaces such as "after 75 min. of training 40 typical users should be able to accomplish 80% of the benchmark tasks in 35 min. with fewer than 12 errors." A professional approach to user interface design puts the responsibility with the designer, having the user or customer specify performance requirements rather than details, just as in Shneiderman's hypothetical acceptance test. This leaves the design of the interface to the software team, but with explicit standards for the specification of the performance of the system.

A second consequence is that we must expect to have to go through many iterations on details of the interface. If the overall project is to be subdivided by dividing the system into modules that are separately designed and implemented, then a third consequence is that the user interface should be segregated into a single module which all other modules must use if they communicate with the end user.

Separating interface design from program design. This strategy consists of isolating the main program in one set of modules, the user interface in another, and considering the user as comprising a third:

```
Interface  Program
  User <---|<--- Module  Modules
          |          |
          |          |  The virtue of this view is well known in the software engineering community: as long as the communication structure and the data representation are well specified and adhered to, the modules can be worked on independently and changed at will, without any effect on the rest of the system. The communication between the Interface and the User can be changed in any arbitrary manner as long as the communication between the Program and the Interface is not affected. This requires, of course, that the only part of the system that may interact directly with the user is the Interface Module. A well-defined protocol between the main program and the interface module can be defined at the outset, leaving the interface designer free to change the interface without interfering with the independent development of the main program. An important benefit of this modularization strategy is that it allows separate specialists to work on the main program and on the interface module.
```

A good example of this modularization combined with support for a heavily iterative approach to developing the user interface is provided by the database and interface packages TrollUSE (the database)11 and RAPIDUSE (the interface package)12. TrollUSE is a relational database—the "main program" module. RAPIDUSE is a tool for the design of interactive dialogues and systems, allowing for rapid modification and experimentation of the interface (through a transition network specification), yet maintaining constant the necessary communication protocol to the database, TrollUSE. Clearly the design of a database requires different skills than designing a user dialogue, and in fact RAPIDUSE is designed to allow non-programmers to design the dialogue using a simple interpreted language. Other examples are the DMS system11 and the work by Burton et al.14.

Makeup of interface design teams. A further consequence of the special nature of interface design affects the organization of the design team. If the principles of interface design were developed to the point where their application required no personal expertise, and if the tradeoffs were all identified and quantified in advance, then a split between groups of experts might be workable. It is noticeable that some of the more successful interface designs (e.g., the Xerox Star and the Apple Lisa) have been done by teams which included people with diverse talents and training, but which did not make a distinction between evaluators and implementors, between psychologists and programmers. We interpret this to indicate that at the present time the closest cooperation is required to identify and resolve the unexpected tradeoffs that surface in the course of a design. Many design questions are at heart unforeseen tradeoff decisions, and these can only be reasonably made by people who appreciate all the factors involved. This, although a separation of the interface from other program modules is highly beneficial, and although logically the skills required for designing each are quite different (e.g., dialogue design versus application programming), it seems that the unknowns in interface design continue to demand close cooperation between parts of design teams.
Influences of the Interface on Code

Languages for User Interface Design

We have argued that interface designers need to be aware that they are designing a dialogue, and more than that a co-routine, between user and computer. The idea immediately follows that special languages that represent this view would be an important aid. Certainly conventional languages are poor in this respect because they are machine-centered: they describe events around the sequence of machine actions with input and output seen as side-effects. Wasserman's RAPID/USE system offers the designer a language based on Transition Diagrams, where nodes correspond to a machine state and the output displayed at that point, and the various arcs from a node correspond to alternative user inputs. This is a major step in the right direction. It is perhaps not a complete solution because it is essentially a representation of the machine the user sees (as a quasi-finite state machine) rather than of the co-routine. The user's processes are only implicit (though much easier to see than in ordinary languages) and machine actions are represented only as side-effects of transitions. Until further advances in this area are made, exercises such as flow-charting the user's decision-making and actions will remain important.

The Interface to the User Module

We have argued that the user should be viewed as a module of the system. Good programming practices now demand that one establish a uniform communication protocol early in the design phase and stick to it. Part of the rationale is economy—if module A has to use one protocol for interacting with module B and another for module C, it will take more code to do so. (Or, by analogy, if you think of module A as being the User, then the user has more to learn.) In addition, there is scope for confusion and error in the programming if there is not a uniform protocol (in the analogy, the lack of consistency makes it harder for the user to learn and to apply the protocols).

These arguments apply best to low-level user protocols. The more programs use a given protocol for interaction, the more benefit the user gains from the associated set of skills in using it (i.e., from their associated "subroutines"). Thus a user is helped if all parts of the system use a small number of common protocols for interaction, protocols that the designer can think of as corresponding to subroutines in the user's mind. 7

It is not sufficient to think of the user as a simple system module with a single protocol for interacting with other modules. There are many layers of protocol, just as there are in computer networking. At least three levels may be distinguished: (1) the low-level protocols just discussed; (2) the conceptual model of the domain the designer wishes to present to the user; and (3) the highest level where the user has several concurrent goals (e.g., to send a letter or get a budget analysis) and the computer is one means to the ends. Most of our knowledge about human-computer interaction is at the first two levels: little is understood/know about the highest level of user goals. The Xerox Star, the Apple Lisa and Visible can be seen as paradigm examples of the second level in that the system image was decided early on as a major design decision which defined the context in which the rest of the interface was developed.

Standardised Packages

A major development tool would be the creation of various packages for doing standard interface operations. There are two separate motivations for this:

- To provide a language at the right level for the designer (where the operations are elements of user interaction and lower-level details can be hidden);
- To provide standard modes of interaction across the system to help the user.

The first provision helps the designer to work on the design of the interface undistracted by implementation details. The second provision gives a standardization (consistency) across applications that helps the user, as argued above. Simply providing packages is often enough to get standardization—it makes it easier for a programmer to conform than to dissent. There is great need for software tools for interfaces, including screen management, user-program dialogue, and packages for doing help, argument parsing, history, and undo. A good example is Perlman's 17 general interface package.

Documentation as an Integral Part of the Interface

An obvious consequence of integrating user interface design into the overall software engineering is to integrate documentation and code generation. Mashey and his colleagues at Bell Labs have taken a major step in this direction by using the same file control system for both source code and documentation 16, thus promoting their simultaneous development and allowing immediate checks on whether one is out of date relative to the other. Knuth's recent work is along the same lines. 17 This is most likely to benefit other programmers (rather than users) who will have to maintain the code, since they are often the major beneficiaries of complete and correct documentation. However, if it is true that no interface design is likely to remain fixed for long, then it is not enough for it to be friendly to the user—it must be friendly to its maintainers as well.

The term "documentation" should not be viewed too narrowly. Users get information from a number of sources including manuals, tutorials, error messages, and normal

---

1 A major problem here is to ensure a match between the actions experienced of the user and the capabilities of that user. This is one of the major themes of research on human-computer interaction, and is a non-trivial problem. In general, we would submit that here is where the software designer must interact with a human-engineer of psychology software designer — during the design phase — to develop the specifications of the user actions.
Software Engineering for User Interfaces

We believe that the system specifications should include a statement about the class of user and the kind of training that is to be expected. The system should then be evaluated with that very same class of user, with the same training procedure. If the user then has problems, there is a bug in the system. The bug could be in the training, or in the interface. The point is that we cannot determine just where the problem lies until we have explicit specifications for all aspects of the computer system, including the interface and the user performance. When we have specifications that cover the user, then we can determine how reasonable they are on the basis of the user's abilities. Only when we have determined that the specifications are indeed reasonable can we then claim that the system that fails to meet those specifications is at fault. This lesson applies to all parts of the system, of course, but its implications for assessing the role of the user as a part of the overall system operation seem not to be properly recognized.

Finding Bugs

The only way to find bugs is to test. This means that the system must be put through its paces with the human user, much as programs are put through their paces with test sets of data. Just as a program needs testing by data that exercise every branch of the code, so the user-program interaction needs testing by exercising each possible "branch" of the interaction. Unfortunately, test procedures for user interfaces do not exist.

Note that the testing phase is not apt to be easy. It requires the development of good test problems, of a good pool of users upon whom the tests will be run, and careful observation and evaluation of the result. It is critical that the users upon whom the system be tested reflect the actual user population for whom the system is designed. Psychology has amassed a number of methodological tools that can be of use. Other tools, specialized for this particular problem, need to be developed.

Learning how to ask users for information is as big a topic as learning how to extract useful measurements from computers. For instance, consider a faulty error message. If it is so useless that the user cannot understand it and gets stuck, there is often a bias against reporting the consequent failure to carry out the task successfully because the user are apt to feel that the problems are due to their own inadequacies. On the other hand, if the message is wrong or silly in some way but nevertheless the users succeed in diagnosing the real problem fairly quickly, then they are likely to express their irritation. Note that this means that the non-fatal inadequacy is likely to receive a much higher rate of spontaneous complaints than the much more serious case which causes users to fail completely. Obviously we need to learn how to work around phenomena like this. For instance, using exhaustive checklists in questionnaires ensures that one solicits opinions on all parts of an interface, and, to some extent, allows one to see things such as mass avoidance of a command that no-one complains about spontaneously.

displays. One test of the adequacy of the overall documentation is to introduce an error in the operation of the system while observing a "typical" user in a "typical" situation. The observation of interest is to determine how the user copes with the situation: the design fails the test if the user cannot recover gracefully. It is important to note how the user determines the state of the system and the options that are available, and also to observe what the user actually does (which may be quite different from what the designer had in mind).

The success of the system will usually depend on a combination of information sources and is neither a property of the code alone nor of the documentation alone. The theme is that in order to provide good documentation from the user's point of view it is necessary to identify what information the user needs, and when. Then, it is necessary to provide a channel from the situation to the information. It is relatively unimportant what media are used for this in any given case: they could be messages generated by the system or notices stuck to the terminal. What matters is whether the user is able to get the information that is required. Note that it is not relevant that the information is available in principle. What matters is whether real users, in real situations, can get the answers. If the user cannot solve the problem, then it is the system design that is at fault, whether or not the designer can demonstrate that the relevant information was available to the user: the critical test is the practical one—do real users succeed at the task?

Debugging the Interface

When a piece of software has been implemented it needs to be tested and debugged: the same applies to the user interface. In the past, debugging the interface has generally been left to the customer, whose complaints are classified as changes to the specifications. The net effect tends to be that changes are slow, expensive, resented by the programming team, and do not benefit from any kind of systematic or professional approach. Clearly the field is more than ready for improved practices.

What Is a Bug?

The field of debugging involves many issues. One problem is to determine what counts as a bug, another is to determine what symptoms can be detected in practice (and what proportion of bugs escape because they produce no clear symptoms), and yet another to determine the cause from the symptom. The concept of "bug" is clearly useful in both traditional and interface software engineering, but nevertheless it has no clear definition. Some bugs are clear—if an explicit specification is not met, the implementation has a bug. However there can be bugs in the specifications themselves, and bugs relative to implicit specifications. A crucial part of developing interface engineering will be developing standards that become implicit specifications for all interface programming. (The analogous points for bugs in programs are discussed in Johnson, Draper, & Soloway.)
Deblirin usage completion immediacy thoughts on the user's experience, one could infer it. It is not necessarily appropriate to do so, by merely developing suitable error criteria, the designer knows who has debugged programs with an understanding of the situation. This is not a trivial instance: users who feel that a system reveals their inadequacies will not be anxious to use the system and will resist its introduction into the workplace. Thus, the system will not get used (or morale may suffer). The problem is to devise techniques that allow the designer to realize the nature of the difficulty. It will take extreme sensitivity on the part of the tester to overcome these problems. It is here that the skills of the experimental psychologist are probably essential.

**Debugging Tools**

The use of questionnaires is analogous to a post-mortem in that they are applied after the program has run. One of the most pressing needs for interface debugging is to have facilities analogous to run-time tests built into all computing environments that cause program exceptions for bad addresses, floating point overflow, etc. Although an important function for these is, of course, the protection of other users, they are also valuable for debugging because they stop execution at the earliest sign of trouble and give the programmer a chance to gather information on the state of the process at that point. Ease of debugging is crucially affected by the immediacy of error detection, as anyone knows who has debugged programs with and without array bounds checking. Applied to interface debugging, this means developing suitable error criteria, and then acting on it. It is not necessarily appropriate to halt a program when an error in interface interaction is detected, but at the least, one could create a relevant "dump"—a trace of the whole interaction together with as much information as possible on the user's experience, and their current goals and thoughts at the time of the difficulty.

The various existing techniques for getting at the interaction between the user and the system differ in their immediacy and the information provided. Further removed from the actual interaction is the collection of opinions after some amount of experience with the system (e.g., at completion of the training period). Closer to the actual usage is the use of on-line complaint facilities. A still more immediate record is provided by history traces or driblet files that provide a detailed record of the low-level actions of the user, but without any of the goals or intentions.

Intensions and goals can be gotten by the collection of real-time, thinking-aloud protocols from users while they interact with the system. Each technique offers a different perspective on the interaction.

**Testing the Interface**

**Improving Measures of Performance**

In addition to debugging, a programmer will typically be concerned with examining and optimizing certain measures. The best parallel here is with the problem of improving a program's speed of execution. The conventional wisdom on the timing problem is that a typical program spends 90% of its time in 10% of its code, so the strategy is to identify that 10% and work on tuning it, since work on improving the other 90% will show little effect overall. Thus, profiling tools that show where a program spends its time are important. In improving an interface, several issues are relevant: how many users find a given command problematic?, how problematic do they find it?, how often does the issue arise? As with debugging, we see here a gap between what can be easily and directly measured and the underlying concern of the designer.

Like profiling tools, then, interface tools should produce measures of those things that can be used by the designer to pick the next point of attack, together with a measure of how important it is to do any further improvements. Also like profiling tools, there will be issues of how accurate these measurements are (resolution difficulties) and how representative of the real situation. Ultimately a lot will also depend on the experience and judgment of the designer using the tool. Thus not only do the tools need to be developed, but it will then take a further significant amount of time to accumulate experience in the use of these measures.

Another tool is on-line command usage measurements. It is relatively easy to collect a running record of command use for the various users of a system, thus providing reliable measures of how often commands are used, and by what percentage of users. The frequency of use is important in weighing the priority to be given to problems.

**Benchmarks and Acceptance Tests**

Earlier, we discussed Schneiderman's[2] suggestion that the specification of the interface be given in terms of the acceptance tests to which it will be subjected. This idea can have far-reaching effect in focusing designers' attention on a definite goal for the interface. Whether or not success at a particular test turns out to be a good guide to the user's long-term satisfaction with the product, it is at least at an appropriate level of specification; this is a crucial step in

---

11 There is a major problem of invasion of privacy. It is not appropriate to keep records on the details of individual users' interactions with a system. Our solution is to encode the user's identity so that although the user identity cannot be determined, we can still match up the particular command sequences and program usage with the user code. This is essential in allowing the discovery of common patterns of operation.
User-interface benchmarks will be most clearly useful when the aspects of performance and the situation in which it is to be measured are clearly defined. As a general method by which to judge a whole system, benchmarks are obviously limited; systems differ on many dimensions and benchmarks often generate only a single measure. The use of benchmarks for interfaces is further problematic in these early days since we do not yet know all the crucial variables. For instance, discussions about which of two operating systems are more effective for a class of users are sometimes carried on without considering the communication rate of the channel to the user, yet this crucially affects how much feedback is perceived as a painfully time-wasting nuisance. In general, factors not directly under the control of the engineer may have a dominating effect. Until we are more confident of being conscious of the factors that have a major influence on the measures we are interested in, we will not know how to run benchmarks in which they are held constant.

---

A statistical problem arises in benchmark tests with users that does not normally arise with hardware: unlike computer hardware one can neither get identical people (so that single measurements generalize reliably) nor run a test twice on the same people with identical results (because of learning effects). Even when we understand the causes of variation well enough to apply statistics with confidence, this will still mean running large numbers of trials where one would have been sufficient to benchmark a machine.
Cognitive Science ONR Technical Report List


<table>
<thead>
<tr>
<th>Reference</th>
<th>Author(s)</th>
<th>Title</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>8202</td>
<td>Jonathan T. Grudin</td>
<td><em>Central Control of Timing in Skilled Typing.</em></td>
<td>February 1982</td>
</tr>
<tr>
<td>8203</td>
<td>Amy Geoffroy and Donald A. Norman</td>
<td><em>Ease of Tapping the Fingers in a Sequence Depends on the Mental Encoding.</em></td>
<td>March 1982</td>
</tr>
<tr>
<td>8204</td>
<td>LNR Research Group</td>
<td><em>Studies of Typing from the LNR Research Group: The role of context, differences in skill level, errors, hand movements, and a computer simulation.</em></td>
<td>May 1982</td>
</tr>
<tr>
<td>8205</td>
<td>Donald A. Norman</td>
<td><em>Five Papers on Human-Machine Interaction.</em></td>
<td>May 1982</td>
</tr>
<tr>
<td>8206</td>
<td>Naomi Miyake</td>
<td><em>Constructive Interaction.</em></td>
<td>June 1982</td>
</tr>
<tr>
<td>8207</td>
<td>Donald R. Gentner</td>
<td><em>The Development of Typewriting Skill.</em></td>
<td>September 1982</td>
</tr>
<tr>
<td>8208</td>
<td>Gary Perlman</td>
<td><em>Natural Artificial Languages: Low-Level Processes.</em></td>
<td>December 1982</td>
</tr>
<tr>
<td>8301</td>
<td>Michael C. Mozer</td>
<td><em>Letter Migration in Word Perception.</em></td>
<td>April 1983</td>
</tr>
<tr>
<td>8302</td>
<td>David E. Rumelhart and Donald A. Norman</td>
<td><em>Representation in Memory.</em></td>
<td>June 1983</td>
</tr>
<tr>
<td>8303</td>
<td>The HMI Project at University of California, San Diego</td>
<td><em>User Centered System Design: Papers for the CHI '83 Conference on Human Factors in Computer Systems.</em></td>
<td>November 1983</td>
</tr>
<tr>
<td>8401</td>
<td>Stephen W. Draper and Donald A. Norman</td>
<td><em>Software Engineering for User Interfaces.</em></td>
<td>January 1984</td>
</tr>
</tbody>
</table>
ICS Technical Report List


Air Force

   Life Sciences Directorate, AFL
   Bolling Air Force Base
   Washington, DC 20332

2. Bryan Dallman
   AFHRL/LBT
   Lowry AFB, CO 80230

3. Dr. Alfred R. Fuley
   AFOSR/RL
   Bolling AFB, DC 20332

4. Dr. Genevieve Haddad
   Program Manager
   Life Sciences Directorate
   AFOSR
   Bolling AFB, DC 20332

5. Dr. John Tangney
   AFOSR/RL
   Bolling AFB, DC 20332

6. Dr. Joseph Yaegeula
   AFHRL/LBT
   Lowry AFB, CO 80230

Department of Defense

12. Defense Technical Information Center
    Cameron Station, Bldg 5
    Alexandria, VA 22314
    Attn: TC

13. Military Assistant for Training and Personnel Technology
    Office of the Under Secretary of Defense for Research & Engineering
    Room 30199, The Pentagon
    Washington, DC 20301

14. Major J. Thorne
    DARPA
    1400 Wilson Blvd.
    Arlington, VA 22209

15. Dr. Robert A. Wiser
    OSBRI (ELS)
    The Pentagon, Room 3D129
    Washington, DC 20301

Military Research and Development

1. Dr. Ricardo A. Butler
   NIE-BPH Bldg. Stop 7
   1200 19th St., NW
   Washington, DC 20006

2. Dr. Susan C. Jenkin
   Learning and Development
   National Institute of Education
   1200 19th Street NW
   Washington, DC 20006

3. Dr. Andrew A. Molnar
   Office of Scientific and Engineering Personnel and Education
   National Science Foundation
   Washington, DC 20550

4. Dr. Everett Palmer
   NASA-Ames Research Center
   Moffett Field, CA 94035

5. Dr. Mary Smith
   C-10, Bell Stop Bldg.
   Los Alamos National Laboratory
   Los Alamos, NM 87545

6. Chief, Psychological Research Branch
   U.S. Coast Guard (G-P-1/2/DRX)
   Washington, DC 20353

7. Dr. Edward C. Weiss
   National Science Foundation
   1800 G Street, NW
   Washington, DC 20550

8. Dr. Frank Vitkovic
   U.S. Office of Education
   NV, Maryland Ave. NW
   Washington, DC 20202

9. Dr. Joseph L. Young, Director
   Memory and Cognitive Processes
   National Science Foundation
   Washington, DC 20550

Civilians for Agency
Private Sector

1 Dr. Keith T. Wescourt
Parasystems, Inc.
505 Middlefield Road, Suite 140
Menlo Park, CA 94025

1 William R. Whitten
Hall Laboratories
20-610
Holmdel, NJ 07733

1 Dr. Thomas Wickens
Department of Psychology
Francis Hall
University of California
405 Hilgard Avenue
Los Angeles, CA 90024

1 Dr. Mike Williams
Intralics Inc.
120 University Avenue
Palo Alto, CA 94301

1 Dr. Joseph Wahl
Alphatech, Inc.
2 Burlington Executive Center
111 Middlesex Turnpike
Burlington, MA 01803