RC 17342 (#76618) 10/30/91
Computer Science 6 pages

Research Report

Finger Force Precision for Computer Pointing

Ted Selker and Joseph D. Rutledge

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents and will be distributed outside of IBM up to one year after the date indicated at the top of this page. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties).

IBM Research Division
Almaden • T.J. Watson • Tokyo • Zurich

Finger Force Precision for Computer Pointing

Ted Selker and Joseph D. Rutledge

IBM Reseach Division
T.J. Watson Research Center
Yorktown Heights, NY 10598

Abstract: This paper addresses motor control constraints which affect analog pointing devices used with computer interfaces. We have investigated the accuracy and precision with which subjects can apply force to a small isometric joystick with a fingertip or finger and thumb, for one and two dimensional force specification.

We find the information content of a force application to be in the range of 4 to 6 bits per dimension with visual feedback, and without time constraint. The force application task in two dimensions is no more difficult than in one dimensions and gives twice the information content. Use of an opposing thumb and finger gives no improvement over a single finger. Both inaccuracy and imprecision are concentrated along the direction of the specified force - subjects tend to be both more accurate and more precise in the direction of force than in its magnitude.

Implications for analog pointing devices are discussed.

Finger Force Precision for Computer Pointing

Ted Selker   Joseph D. Rutledge

IBM T.J.Watson Research Center, Yorktown N.Y. 10598

ABSTRACT
This paper addresses motor control constraints which affect analog pointing devices used with computer interfaces. We have investigated the accuracy and precision with which subjects can apply force to a small isometric joystick with a fingertip or finger and thumb, for one and two dimensional force specification.

We find the information content of a force application to be in the range of 4 to 6 bits per dimension with visual feedback, and without time constraint. The force application task in two dimensions is no more difficult than in one dimension and gives twice the information content. Use of an opposing thumb and finger gives no improvement over a single finger. Both inaccuracy and imprecision are concentrated along the direction of the specified force - subjects tend to be both more accurate and more precise in the direction of force than in its magnitude.

Implications for analog pointing devices are discussed.

KEYWORDS
Mouse, Joystick, Dexterity, Force Precision

INTRODUCTION
Recent work [5, 7] has demonstrated that pointing performance can be significantly improved when human motor-perceptual limitations are taken into account.

This reopens questions of strategies for physical control posed at the time of radio knob design, etc. [2-4, 8] What are the perceptual motor constraints of physical control design? How should these constraints affect the relationship between the physical control and a machine's response to its movement? By understanding the relationships between these information channels we can improve design interfaces.

Motivated by work in designing a finger pressure controlled pointer, we wanted to understand the control channels available to various analog motor tasks. We report here a study of the finger-pressure channel. We have considered visual, auditory, tactile and proprioceptive feedback. Some tactile feed back is inevitably present (barring anesthesia). Preliminary experiments indicated that auditory feedback alone is much less effective than visual, even in the one-dimensional case, while tactile feedback alone appears to allow less than 3 bits in each dimension. This study focuses on finger pressure with tactile and visual feedback, without time constraint. The subject may take several seconds to apply the specified force, and the force is then measured as the mean of instantaneous forces sampled over a 2.4 second integrating period. Even under these conditions accuracy and precision are surprisingly low.

The force range which we have investigated is that appropriate for one or two fingers on a small sensor (a 3 mm by 8 mm cylinder), 0 - 225 grams.

While there are many studies of complex tasks such as pointing and tracking which use the finger or hand force channel, there seem to be few if any which address the accuracy and precision available in the channel itself. We have found only [6], which studies whole arm movements at much greater forces. His studies found subjects could apply a arm force to within 10percent of a attempted target force.

METHOD

Apparatus
Subjects were seated at a standard office desk, in a chair adjusted to comfortable height by the subject. On the desk were a CRT display and a 101-key IBM PS/2 keyboard, placed about 10 cm back from the edge of the desk. In the center of the keyboard, between the G and H keys, was an isometric joystick (the same sensor used in the Pointing Stick, [5]) topped by a dished 3x5 mm diameter finger rest, 4 mm above the level of the key caps Figure 1. The joystick top moves an undetectable .13 mm at maximum force.

Figure 1. Fingertip grip of an isometric joystick

An IBM 8514 in VGA graphics mode .4 mm square pixels on a 480 by 640 screen

For the fingertip grip condition the subject placed a finger tip on the finger rest. For pen grip experiments, adjoining key caps were removed exposing an 8 mm long 3 mm diameter "pen", grasped between thumb and forefinger. Figure 2

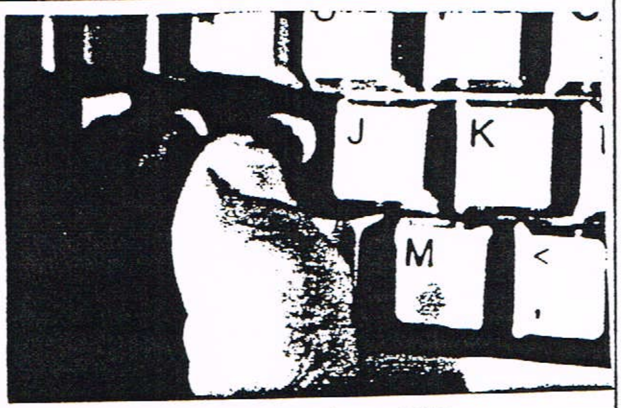

Figure 2. Pen style grasp of an isometric joystick

Experiments were also run using a single finger or thumb pressure on the side of the joystick to understand the value of the opposing digit in the pen grip condition. This is called the side grip.

Strain gauge signals from the sensor were processed to produce signals on an IBM PS/2 pointing device interface such that the resulting cursor position represented the horizontal force being applied to the sensor, within the limits of the display screen. Strain gauge sensing and signal processing was performed by a separate IBM PC/XT with a Scientific Systems Labmaster data acquisition board which communicated to the PS/2 through its mouse port.

A program running on a IBM PS/2 model 80 presented stimuli, provided feedback, and recorded data.

Subjects
Over several month periods one subject performed experiments to calibrate and develop data collection techniques.

We report here on results from four subjects, hired through an agency as office temporaries. They all normally work in secretarial and clerical jobs, frequently using word processors, and had slight familiarity with mice, but no prior experience with other computer pointing devices. All were women, between 25 and "50's", who reported typing speeds between 50 to 80 words per minute. Two played or had played a musical instrument; no other high-manual-dexterity hobbies were reported. Subjects participated in these experiments as part of a two-day sequence of experiments on pointing behavior, using the Pointing Stick in its normal mode and a mouse in addition to the present apparatus.

Procedure
The experimental paradigm is as follows: Subject initiates each trial with a keypress. A target force is presented as a position on the screen. The subject attempts to apply the specified force, by bringing the cursor to the target and holding it there.

An initial movement towards the target is ended when the subject's precision limit is inevitably reached and the cursor moves away from the target. During the following "hold" phase the cursor is held as stably as possible for some 2.4 seconds. The mean applied force during the "hold" phase will be called the trial's **effective force** and its standard deviation is taken as the trial's imprecision or **dither**. The **miss vector** is the difference between the target vector and the effective force vector. The **miss angle** is the direction of the miss vector.

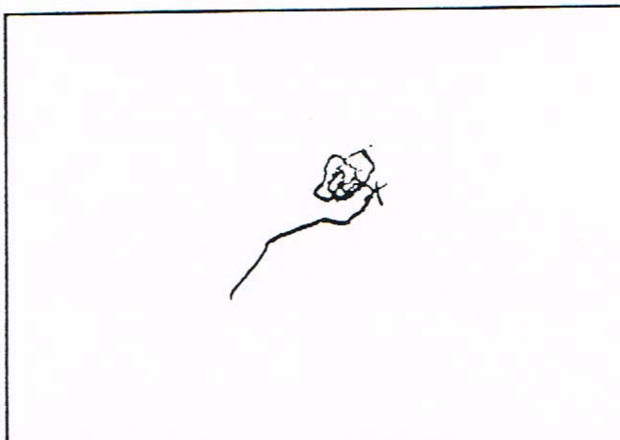
Figure 3. Display showing stability and accuracy of force application in a trial.

Feedback information is added to the target display at the end of each trial. The computer displays the track of the cursor's movement. A cross on the track marks the beginning of the hold phase and an ellipse is centered at the mean hit position with shape representing the variation in applied force during the hold phase.

Separating subjects target selections from target hold phase is exemplified in Figure 3. In this figure it is noticeable that after a subject attempts to select a target, stability at holding the requested force can be characterized as a cloud of dither.

After 10 such trials the computer displayed numbers indicating relative accuracy and relative dither and standard deviations for the group of trials. A menu selection is selected to run ten groups of these.

The subjects performed these the trial type doing two dimensional (circle) targets, vertical one dimensional, and horizontal one dimensional targets. All of these conditions were investigated for the finger tip as well as pen grip conditions. A nominal 300 trials were collected for each two dimensional condition and a nominal 100 were collected for each one dimensional condition. For two subjects an additional nominal 100 trials were collected for the side grip condition.

Target forces were pseudo-random, with uniform distribution over the range corresponding to the display screen, excluding an approximately 1 cm margin and a disk of diameter 2 cm in the center.

An experimenter remained in the room to observe and to demonstrate protocols. Written instructions explained each phase of the experiment, and directed the sequence of phases; the sequence was varied in some cases to maintain subject motivation. The experimental protocol was otherwise administered and results recorded by the computer.

RESULTS

Target force, mean applied force, dither, and time were recorded for each trial. Initial sets of trials were discarded, as were a few trials invalidated by interruptions or equipment problems. Table I gives overall averages of miss and dither for the three target types, and for the two grips. Figures 4 - 12 look at the data in more detail, examining dependencies on target force,

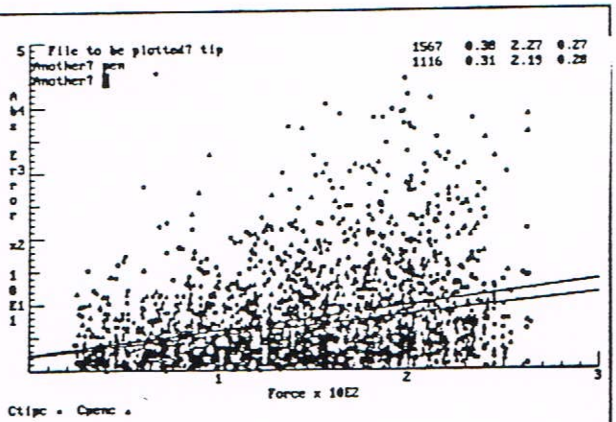

Figure 4. The relationship between distance and accuracy

Figure 4 plots error against target force for the two grip conditions. The figure shows that while absolute accuracy of force decreases with requested force, relative accuracy increases with force. The large range in accuracy indicated by the quite tall cloud shows that range in performance is wide. Note that the pen grip condition adds a slight advantage over the finger tip condition.

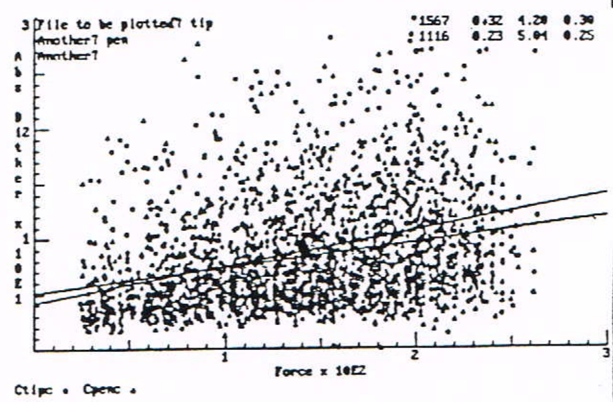

Figure 5. The relationship between force and steadiness

Figure 5 plots dither against force for the two grip conditions. Notice the blank area at the bottom of the screen. This represents a limit to steadiness. The width of this band shows that no one was able to hold force steady to within 2 grams. The average instability was 8 grams varying in individual from 5 to 12 grams. Like the error, dither also increased somewhat with force.

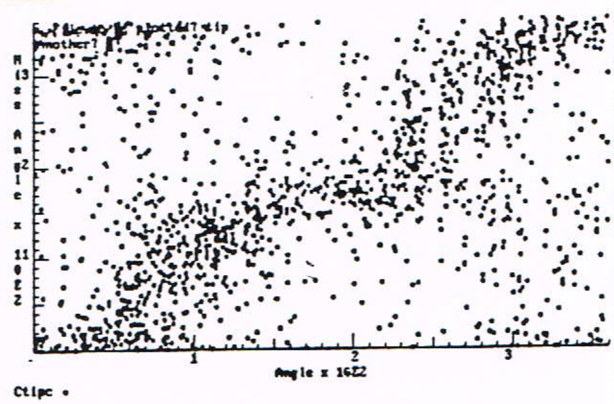

Figure 6. Direction of miss versus target direction

Figure 6 shows the relationship between the target direction and the miss angle. Note that the miss angle tends to be 180 degrees from the target angle showing that subjects tend to undershoot the target force. It shows that subjects were more accurate in the direction than in magnitude of force application. This data could be used to help design velocity transfer functions. Another use of this suggests that menus should be made "deep" in the direction of most likely approach. Edge menus, which are effectively infinitely deep, are instances of this.

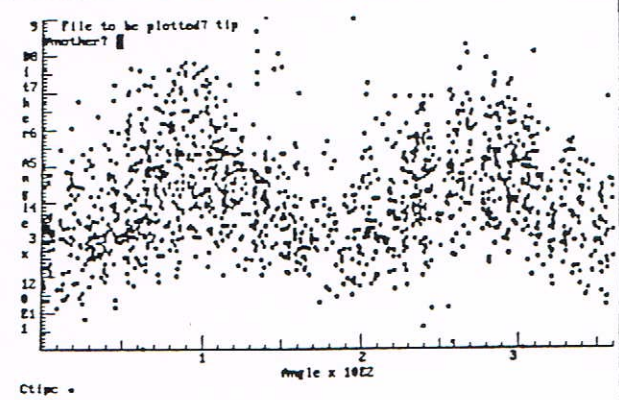

Figure 7. Stabilty in direction of target is worse than in orthogonal direction

A striking result emerges by looking at the shape of the data cloud of subject hold data, Figure 7. The ellipse that represents the mean of the dither cloud has approximately a 2 to 1 axis ratio in the direction of the target. We have collected data for pressure towards targets above below to the right and to the left. Our data collection procedure for angular forces does not distinguish this at other angles.

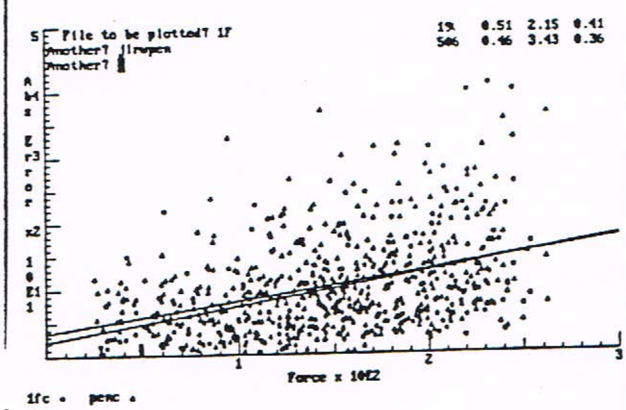

Figure 8. Demonstration that pen grip is not better than pushing with a side grip.

Experiments comparing the side grip to the pen grip test the role of opposing grip for pressure accuracy and stability Figure 8. Note that in the data the side grip is not distinguished from the pen grip. This suggests that the slight advantage of the pen grip condition over the tip grip is due to improvement in control by pushing on the side of the post rather than to any additional stability provided by another finger.

	Samples	Error	Dither
2 Dimensional			
Tip	1567	7.7	8.8
σ		7.8	6.0
Pen	1116	6.6	8.3
σ		6.1	5.2
1 Dimensional - Horizontal Bars			
Tip	389	4.5	6.0
σ		4.7	4.0
Pen	260	3.3	4.8
σ		4.3	4.2
Horizontal Projection of 2 Dimensional Data			
Tip	1510	5.3	6.2
σ		5.7	4.2
Pen	1074	4.7	6.2
σ		4.9	4.1
1 Dimensional - Vertical Bars			
Tip	390	5.1	5.6
σ		7.0	5.2
Pen	397	4.3	5.6
σ		4.9	4.4
Vertical Projection of 2 Dimensional Data			
Tip	1519	4.6	6.0
σ		6.4	5.0
Pen	1079	3.9	5.4
σ		4.7	3.9

Figure 9. A Table of average error and dither in grams showing the inherent low accuracy and instability in force application, the relation between 1 and 2 dimensional trials and the improvements afforded by the pen grip.

	Samples	Error	Dither
Subject 1	293	3.0	5.8
Subject 2	587	4.5	6.9
Subject 3	330	15.0	10.6
Subject 4	357	9.9	12.7

Figure 10. A Table showing the range of individual differences in average error and dither force between subjects for the two dimensional tip grip condition

Figure 9 gives the average error and dither under the various conditions. This table shows that even when the data is averaged over all individuals and target forces its major features are still visible. The extremely large standard deviations reflect several features of the data: first that subjects are unable to hold a constant force, second the large variation in accuracy between individuals Figure 10, and third the fact that error and dither increase with increasing force.

The maximum force that subjects seem comfortable applying with one finger is about 8 ounces or 225 grams. Taking this as the working range for finger pressure, average short-term precision can be calculated from Figure 9 The smallest interval .of error for the 2 dimensional tip case is 8 grams. Adding the standard deviations to error and dither for the 2 dimensional tip case gives a value of (7.7 + 7.8) + (8.8 + 6.0) = 30 grams as representing an upper range for error. This gives a precision of 3 to 12 percent, allowing 8 to 30 choices in each dimension Figure 9 on page 4. Taking the log base two of these numbers gives an information content ranging from 3 to 4.9 bits for each trial. Averaged over a long hold time (2.4 seconds) subject accuracy came to a precision of 3 to 15 grams, giving 3.9 to 6.2 bits, for each dimension, Taking a long careful attempt, then, only allowed our best subject to be able to accurately select one of 70 force values in each dimension. Compiled into two dimensions this gives a total information content of 7.8 to 12.2 bits.

The general level of information content reported by Schmidt et.al. for arm force [6] (Standard deviation about 10 percent of applied force) is consistent with our results.

The 1 dimensional projection of the 2 dimensional horizontal data results from setting all horizontal components to 0 in the data from 2 dimensional trials. Note that it agrees closely with the data for the one dimensional targets. Data from trials with vertical targets and the corresponding projection of the 2 dimensional trial data are quite similar, with the same close correspondence. The differences in error between 1 and 2 dimensional trials is completely accounted for by the dimension of the trail. The subjects appear to be able to add a second dimension to their task without any interference giving twice the information content without additional effort.

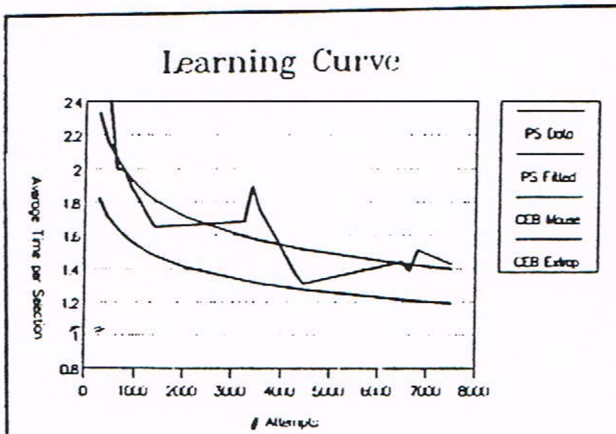

Figure 11. Demonstration of the experience subjects had relative to asymptotic improvement with use of pointing device.

The experiments described in this paper were included in a two day exercise designed to train subjects in using isometric pointing devices for mouse-like selection activities. As with the mouse [1] we find that a subject improves at selection tasks over a few thousand pointing selections. Figure Figure 11shows two points at which Subjects performed these force experiments as part of a longer set of runs designed to give them experience with isometric pointing.

By noting the points along the performance curve at which data was gathered for force experiments, the impact of pointing experience on these results can be estimated. Figure Figure 12 shows that this data is consistent with all other data describing accuracy of finger force.

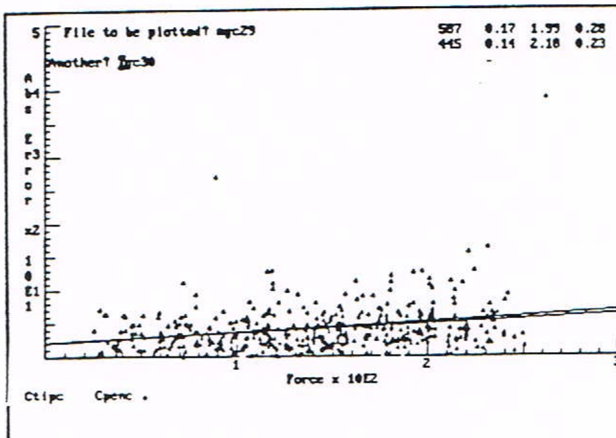

Figure 12. Force accuracy data taken at different times on a user's joystick pointing learning curve.

CONCLUSION

Limits of Digit Force Control

Even under the optimal condition of long integration time, accuracy is quite limited. If we take the working range for this kind of finger pressure as 0 - 225gm, average short-term precision, for our subjects, is in the range of 9 to 25 grams or 4 to 10 percent of full range in each dimension, corresponding to an information content of 3.3 to 4.6 bits for each trial. Over time (2.4 seconds) this is integrated to a precision of 3 to 15 grams, giving 3.9 to 6.2 bits, for each dimension, for a total information content of 7.8 to 12.2 bits. This compares with 18.22 bits represented by the selection of a single pixel on a VGA (640x480) screen. A force-to-position joystick is clearly not adequate as a pointing device. By mapping force into velocity instead of position, efficient time integration and arbitrarily precise pointing may be achieved, provided that the proper mapping is used. The imprecision of finger pressure implies that speed can be closely controlled only where the mapping has a plateau, where the desired speed is maintained over a range of pressures larger than the 4 to 10 percent uncertainty. In the Pointing stick transfer function such plateaus are found at zero speed for a stopped cursor, a slow speed for accurate pixel and character positioning, and at maximum eye tracking speed for fast accurate movements [5] -- Figure id 'transfe' unknown --

Over the range of target forces tested subjects were never able to keep steady forces under 2 grams. A projected average 5 grams of dither at zero force sets a minimum deadband which will allow a user to rest a finger on a sensor without cursor motion.

Two Dimensions is No Harder than One.

Subjects applying a force in two dimensions made much the same errors along each axis as when they were concentrating on that axis alone. One might expect that it would be easier to apply and hold a specified force to the right, for example, if one need not control in the up or down direction. This is not the case.

Role of Grip in Force Control

It might have been expected that the opposing thumb and forefinger hold would greatly improve both accuracy and steadiness of force application. Data from this study shows that the pen grip improved performance very little (Figure 4 on page 3, Figure 9 on page 4). This improvement, however, does not come from the opposing thumb as predicted; a finger placed on the side of the joystick in the side grip condition slightly outperforms the pen grip Figure 8 on page 4shows that the observed force accuracy advantage of the pen grip is due to the finger or thumb position on the side of the sensor rather than on the top, and not to the opposing digits grip. Two unsteady fingers are just as unsteady as one unsteady finger.

Error and Dither are Aligned in the Target Direction

Applied force tends to be aligned to the direction of target force but undershot. For force to velocity transfer functions this is an advantage; the noise along the intended direction of movement changes the cursor speed but not its direction.

A specific style of menu that this suggests is a menu along the edge of a screen (which acts as if infinitely deep).

The results of these experiments help describe behavioral motor issues which contribute to the performance improvements achieved by the Pointing Stick transfer function.

This study shows that individual performance differ by a large factor; could this be taken into account in personalized or adaptive transfer functions?

The noise in a persons movement is greater in the axis directly intersecting the target; could this analysis be utilized to to augment user control?

1. S. K. Card, W. K. English, and B. J. Burr. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys,. *Ergonomics*, 21(8):601-613, 1978.

2. P. M. Fitts. The Information Capacity of the Human Motor System In Controlling T. *Journal of Experimental Psychology*, 47:381-391, 1954.

3. P. M. Fitts and J. R. Peterson. Information Capacity of Discrete Motor Responses. *Journal of Experimental Psychology*, 67(2):103-112, 1964.

4. William Leroy Jenkins and Minna B. Conner. Some Design Factors in Making Settings on a Linear Scale. *Applied Psychology*, 33:7-25, 1949.

5. Joseph D. Rutledge. and Ted Selker. Force to Motion Functions For Pointing. *Interact '90 Proceedings*, North Holland Amsterdam, August 1990.

6. R.A. Schmidt, N.H. Zelasnik, B. Hawkins, J.S. Frank, and J.T. Quinn, Jr. Motor Output Variability, A Theory for the Accuracy of Rapi. *Psychological Review*, 86:415-451, 1979.

7. Andrew Sears and Ben Shneiderman. High Precision Touchscreents: Design Strategies and. Department of Computer Science University Of Maryland., CAR-TR-450. 1988.

8. R. S. Woodworth. Accuracy of Voluntary Movement. *The Psychological Review, Series of Monograph Supplements*, 3(3):1 114, July 1899.

Copies may be requested from:

IBM Thomas J. Watson Research Center
Distribution Services F-11 Stormytown
Post Office Box 218
Yorktown Heights, New York 10598

RC 16910 (#74963) 5/29/91
Computer Science 5 pages

Research Report

Demonstrating Usability Improvements in Automatically Presented Adaptive Help

Ted Selker

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

Kevin M. Goroway

Worcester Polytechnic Institute
Worcester, MA

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents and will be distributed outside of IBM up to one year after the date indicated at the top of this page. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties).

IBM Research Division
Almaden • T.J. Watson • Tokyo • Zurich

Demonstrating Usability Improvements in Automatically Presented Adaptive Help.

Ted Selker
I.B.M. T. J. Watson Research
Yorktown Heights, NY.

Kevin M. Goroway
Worcester Polytechnic Institute
Worcester, MA.

April 15, 1991

Abstract

This paper presents a study demonstrating improvements in performance and perceived usability for users of an automated adaptive help system. COgnitive Adaptive Computer Help (COACH) is an architecture for experimenting with adaptive user models. The language coached on, the way COACH adapts, and the pedagogical paradigm can be changed by writing a language definition table, courseware facts and COACH learning rules. In a five session user study which was disguised as a Lisp course, the COACH system, which gives automatic adaptive help, was found to improve both performance and perceived usability when compared to a version which offered only nonadaptive, user requested help.

Keywords

Models of the user, User interface design issues, Managing human factors in system development, AI in education

Type of submission

Paper

Word count

3687 including Abstract

Demonstrating Usability Improvements in Automatically Presented Adaptive Help.

Ted Selker
I.B.M. T. J. Watson Research
Yorktown Heights, NY.

Kevin M. Goroway
Worcester Polytechnic Institute
Worcester, MA.

Abstract

This paper presents a study demonstrating improvements in performance and perceived usability for users of an automated adaptive help system. COgnitive Adaptive Computer Help (COACH) is an architecture for experimenting with adaptive user models. The language coached on, the way COACH adapts, and the pedagogical paradigm can be changed by writing a language definition table, courseware facts and COACH learning rules. In a five session user study which was disguised as a Lisp course, the COACH system, which gives automatic adaptive help, was found to improve both performance and perceived usability when compared to a version which offered only nonadaptive, user requested help.

1 Introduction

The possibility of real-time interactive adaptive systems has always been questioned[2; 6]. Behavioral studies have not been able to show performance differences for people using adaptive interfaces [5].

We have been working for some years with interfaces which offer users help and tutoring support when needed, without interrupting user tasks [3].

COACH is an adaptive interactive help system which interactively changes its model of a user character by character. It utilizes this adaptive user model to attempt to understand a user's needs. The adaptive system watches a programmer's progress, advising with syntax, examples and descriptions as needed. It ceases advising on a topic in which its user has demonstrated expertise. Help text reappears when the user once again needs help. Whether giving help for user-initiated explanations, or acting as an advisor for COACH-initiated explanations, COACH presents information at the level at which a user has shown competence.

The COACH interface (see figure 1) allows the user to type without computer interruption. To avoid confusion, the Emacs editor window, in the bottom left corner of the screen, is separated from the output text window. The output text window is immediately to the right of the editor window. These two windows can clearly be seen in figure 1. With this configuration the result of the user's input text does not clutter their input pane.

A small window, directly above the editor window, presents help relavent to the immediate context. It displays error messages, when appropriate, as well as information or help on the current construct the user is working on. Even when an error is pending the user is not halted from typing. This window displays the error, allows the user to finish their current thought, then use the editor to correct their error. When an error is not pending this window tells the user what the system is expecting, such as an open parenthesis or a function name. Following this is an example of what the system is expecting along with a description and the syntax of the expected item. This help text can be presented at one of four experience levels, ranging from beginner to expert. The level of information that is presented to the user is determined by the COACH adaptive user model. When a user demonstrates knowledge of a particular concept COACH recognizes this and presents help on that concept at an appropriate level. COACH also takes into consideration errors made by a user, and updates its internal information on the user accordingly.

Above the immediate help window is a large general help window. When a user asks for help it is displayed here. The system updates this window whenever the user moves on to a new concept. For example, it might start by explaining what an *evaluable* is, and when the user has typed `(setq` COACH would explain, at the level at which this user has demonstrated expertise, how one uses `setq`.

Along the bottom of the screen are various menus that the user may click on to receive information such as functions that are available, undefined variables that have been referenced, and help on all of the constructs that COACH knows about. Menus are also available when the user hits the right mouse button.

This arrangement was chosen to minimize visual searches. It is our idea that the top most window will be used least during an editing session, whereas the error window will be used more frequently. Window layout can reflect user needs. To keep work context an advanced user needs to see more of their work and less help: they are supplied with a larger editor window and a smaller help window.

This paper describes a study designed to demonstrate the effectiveness of an automated, real-time, interactive, adaptive user interface as compared to the same interface without the automated, adaptive features.

2 Evaluating Adaptation

The current study tests the hypothesis that a coaching paradigm improves user productivity. Does the adaptation of COACH or the automatically presented help make COACH usable? A combination of these two things, or something entirely different (ie. the online mouse selectable help, the separate input and output windows, the automatic real-time error detection, etc.) could make an imwas observed trying

```
((plus 5 5) times (plus 2 2))
```

but was able to understand the problem with the aid of COACH. The system quickly recognized that an error was being made and popped up an attention grabbing error message in the immediate help window. When this error message was noticed the user was able to use the COACH online help to figure out the proper order of evaluation.

The list operations introduced CAR, CDR, and CONS, as well as the concept of nested lists, and QUOTE. The students were asked to create some lists of varying degrees of difficulty. The simplest was a single level list, (1 2 3), and the most difficult involved a multiple level list that required an understanding of quoting.

A conditional statement was required to solve one of the questions. The task was to print "yes" if a certain element was contained in a list, therefore the MEMBER function was also needed.

When the students finished these single answer questions at their own pace they were asked to begin a database project. The students were not told which functions they should use nor how they should construct the database. Surprisingly, most students chose to use DEFSTRUCT instead of lists or property lists, all of which were covered in the Lisp Tutorial. The reason most students gave for this choice was that, "It does everything for you."

3.4 Comment Sheets

Each day the subjects were asked to fill out a comment sheet. They were instructed to write as much, or as little as they chose. The following questions were asked on the comment sheet:

1. How often do you look at the help screen while solving a problem?
2. How helpful is the Help screen?
3. How helpful is the COACH window system, as compared to a line-based interpreted environment?
4. Observations about COACH?
5. Observations about Lisp?
6. What problems are you having?
7. What problem are you working on?
8. What is your motivation to learn Lisp?: 1–10

3.5 Interviews

Near the end of the course, eight subjects from each group were interviewed on audio tape while they were working. The important question that was asked is:

- What do you find most helpful while solving a problem, the help screen, the COACH menus, or the Tutorial?

3.6 Post-Test

At the end of the user study the subjects were given a post test. To measure the amount of Lisp learned by each subject, questions covered the same material as the pre-test. In addition, questions about specific feelings toward COACH and Lisp were posed. Although similar to the comment sheets, these questions were worded differently to reveal more about the subjects' personal views.

4 Method

Nineteen employees of IBM T. J. Watson Research Center were recruited. They varied from summer interns to professional programmers. While all of the subjects had prior programming experience, none had previous experience with Lisp.

The subjects were separated into an early session (meeting from 5:00pm to 6:00pm) and a late session (meeting from 6:00pm to 7:00pm) each day, for five days. Attempts were made to assign subjects to whichever session fit their schedule best. Eight subjects were assigned at random to use the manual help system. Eleven subjects were assigned to use the automatic adaptive help system.

5 Experimental Data and Analysis

The following sections give detailed descriptions and analyses of data taken in the study. The Data recorded came from the pre-test, comment sheets from each day, saved exercise solutions from each student, verbal interviews of subjects during their sessions, and the post-test taken after the course was completed.

5.1 Pre-Test

Before the course began the students were tested on their knowledge of Lisp, and programming concepts in general. The pre-test showed that all subjects were experienced programmers, having no prior Lisp experience. Answers to the Lisp specific questions showed that, except for the simplest cases, the subjects were unable to answer Lisp questions by merely guessing.

5.2 Saved Exercise Solutions

The COACH system internally stores information about each user. We refer to this internal representation as a user model. When a user is finished using the system for a period of time, COACH creates two files. The first contains the users work (Lisp code), and the second contains their user model (What level of expertise they are at for specific concepts).

After looking through the subjects' saved work and their saved user models the main result of the study was determined: *the users of the adaptive system wrote five times as many functions.*

Although the comment sheets all indicated that the students were working on the database project by the last session, their actual work showed a different pattern. The users of the adaptive system wrote an average of 2.5 functions during the course as compared to 0.5 for the users of the nonadaptive system. No user of the nonadaptive system wrote more than two functions. In addition the style and quality of functions written by the adaptive system users was much better than that of the control group.

One user of the adaptive system was able to write:

```
(DEFUN add-person (name phone lang)
  (COND
    ((MEMBER name USERS))
    ((EQUAL USERS nil)
      (SETQ USERS (LIST
```

non-self-paced manner.

7 Future Work

COACH is designed to demonstrate the use of adaptive automated help. This study has shown that an adaptive automatic help system increases performance.

Many important questions remain unanswered. The COACH adaptive user model was used in a simplified form for this study. It is important to examine the various kinds of adaptation and knowledge bases in such an adaptive user interface. A small number of adaptive rules were used to change the quantity and quality of help given to a user for each function, token, and concept. The value of each individual rule should be studied. COACH also has facilities to record user information concerning required and related knowledge.

In this study we tested the effectiveness and importance of an adaptive system with Lisp illiterate users. We believe the way COACH will be helpful to novice programmers as compared to experts is quite different. Experts will benefit from COACH's anomalous examples and complete syntactic descriptions of functions, leaving the expert alone when they are working on something that they are experienced with. Experienced Lisp programmers will benefit from the fact that COACH keeps track of context dependent situations, scope, and undefined variables, while exposing the user to the relationship between functions and concepts. Novices, on the other hand, find the use of the changing adaptive help for functions and tokens quite useful for learning the syntax of simple functions and token types. While studies demonstrating these effects are straightforward, they were beyond the scope of this experiment.

8 Conclusions

In this study significant differences were found in groups which had adaptive automatic COACH help versus subjects who only manual COACH help.

While data collected from the comment sheets from the two groups shows indistinguishable motivation and self described performance between them, the automatic adaptive help group utilized all available materials (the automatic COACH help, the Lisp Tutorial, and user requested COACH help), felt more comfortable with Lisp, had a higher morale and wrote five times as many functions on average than the group with manual COACH help.

While both groups had the same access to the tutorial and online help, the manual help subjects were likely to look only at menu selected help, in contrast to the automatic adaptive help group, which used tutorial materials as well as asking for help. The reason for this is most likely a lower morale among the manual help subjects; they missed the student/teacher interaction of a classroom environment. Although the self-perceived motivation of the manual help group was not much different than the other group, the data clearly shows that the manual help group put forth less effort.

The automatic adaptive help system succeeds in raising morale, and motivation, which in turn leads to an attitude that is more conducive to learning.

Acknowledgments

We would like to thank the Fame Group at T.J. Watson for all of their technical help, Matt Kamerman for his help in the development of COACH, and John Kolassa for his help with the statistical analysis. We would especially like to thank Bob Campbell for his excellent ideas that we used to collect data.

References

[1] K. A. Brownlee. *Statistical Theory and Method in Science and Engineering.* Krieger, Malabar, FL, 1984.

[2] D. R. Gentner. A tutor based on active schemas. *Computational Intelligence*, 2, 1986.

[3] T. Selker. Cognitive adaptive computer help (coach). In *Proceedings of The International Conference on Artificial Intelligence*, pages 25–34. IOS, 1989.

[4] T. Selker and K. Goroway. A study demonstrating usability improvements in automatically presented adaptive help. Technical report, IBM T. J. Watson Research Center, Yorktown Heights, NY, 1990.

[5] N. Walker and J. B. Smelcer. A comparison of selection times from walking and pull–down menus. In *Proceedings of CHI '90*, pages 221–225. Addison Wesley, 1990.

[6] A. Y. Zissos and I.H. Witten. User modeling for a computer coach: A case study. *Int. J. Man-Machine Studies*, (23), 1985.