UMTRI

90522

Technical Report UMTRI-96-37 December, 1997

Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models

Daniel Manes,
Paul Green,
and
David Hunter

UMTRI HUMAN FACTORS

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UMTRI-96-37		
4. Title and Subtitle **Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models**		5. Report Date December, 1997
		6. Performing Organization Code account 032923
7. Author(s) Daniel Manes, Paul Green, and David Hunter		8. Performing Organization Report No. UMTRI-96-37
9. Performing Organization Name and Address The University of Michigan Transportation Research Institute (UMTRI) 2901 Baxter Rd, Ann Arbor, Michigan 48109-2150		10. Work Unit no. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Road Commission of Oakland County (RCOC) 31001 Lahser, Beverly Hills, MI 48025 USA		13. Type of Report and Period Covered 10/84 - 1/97
		14. Sponsoring Agency Code

15. Supplementary Notes

This research was funded by RCOC and the Federal Highway Administration.

Thirty-six drivers entered and retrieved destination using an Ali-Scout navigation computer. Retrieval involved keying in part of the destination name, scrolling through a list of names, or a combination of those methods. Entry required keying in the destination's name, the longitude, and the latitude.

For young subjects, mean interkeystroke intervals were 1.7 s for initial cursor keystrokes, 1.5 s for initial enter, letters, and shift keystrokes, 1.2 s for numbers, and 0.6 s for space (different from the "standard" value in keystroke-level models of 1.2 s for a "worst" typist). Second keystrokes were about 1 s for letters, 0.7 s for cursor actions, and 0.5 s for numbers, similar to the standard time for complex codes (0.75 s/keystroke). For more than 2 cursor keystrokes, times were about 0.5 s. Age differences were large, with middle-aged drivers taking 40 percent longer and older drivers 120 longer than young drivers. Mental (thinking) times averaged 2.2 s, much greater the standard time (1.35 s).

Equations were developed linking keystroke-level predictions to the actual times. Linear equations based on pure keystroke-level models accounted from 41 (all subjects) to 78 percent (young subjects only) of the variance of retrieval times and 12 to 39 percent of entry times. For tailored keystroke-level models (with experimentally-based values for keystroke times, and adjustments for age and lighting) the variance accounted for was 58 and 83 percent (all subjects and young subjects, respectively) for retrieval, and 47 and 49 percent for entry. Use of linear equations and tailored models significantly reduced the size of prediction errors, making the predictions useful for engineering evaluations of alternative interfaces.

17. Key Words	18. Distribution Statement
ITS, human factors, ergonomics, driving, navigation, route guidance, usability, input devices, controls, keyboards, GOMS, keystroke models	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classify. (of this report)	20. Security Classify. (of this page)	21. No. of pages	22. Price
none	none	64	

Form DOT F 1700 7 (8-72) Reproduction of completed page authorized

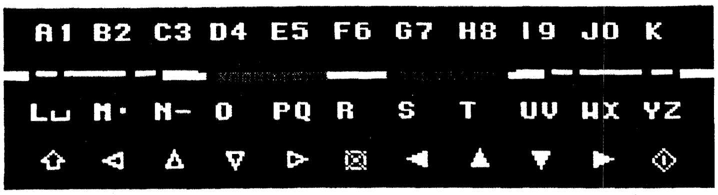
Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models

UMTRI Technical Report 97-42

Daniel Manes, Paul Green, and David Hunter

University of Michigan

Ann Arbor, Michigan, USA



The Ali-Scout Keyboard (95% actual size)

1 ISSUES

1. What are typical times for keystrokes and mental operations for navigation data entry using a compact alphanumeric keyboard with poor tactile feedback?
2. How well do various keystroke-level models predict actual performance?

2 METHOD

# of Subjects

Age	Men	Women
18 - 30	6	6
40 - 55	6	6
> 65	6	6

# of Trials

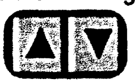
Task	Real Ali-Scout		Simulated Ali-Scout
	Dusk	Night	Dusk
Retrieve destination from unit's memory	5	5	5
Enter new destination	5	5	5

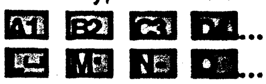
Task #1

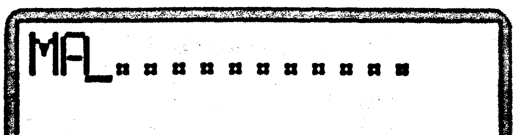
Retrieve destination from memory

- Retrieve MAIN THEATER

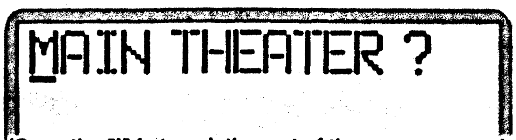
Scroll through list OR Type characters







(Subject has begun typing "MAIN")



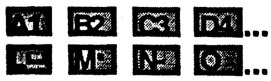
(Once the "I" is typed, the rest of the name appears)

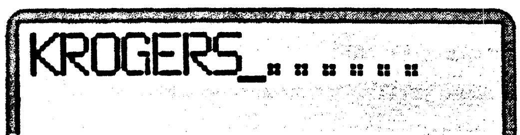
Task #2

Enter new destination into memory

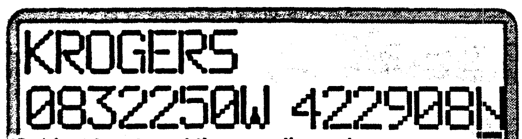
- Enter KROGERS with coordinates (0832250W, 422908N)

Type characters ONLY



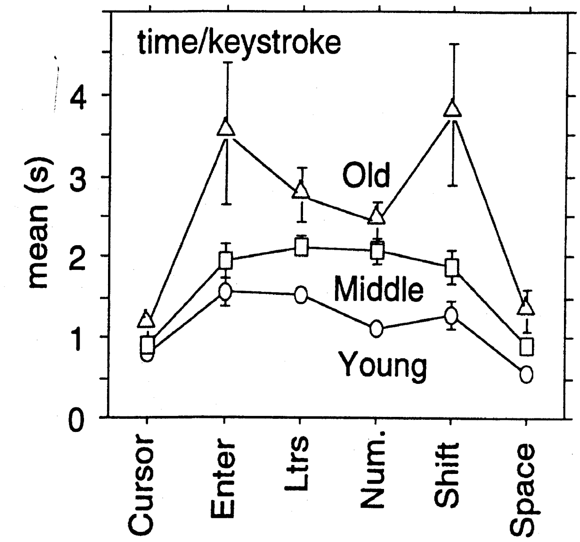


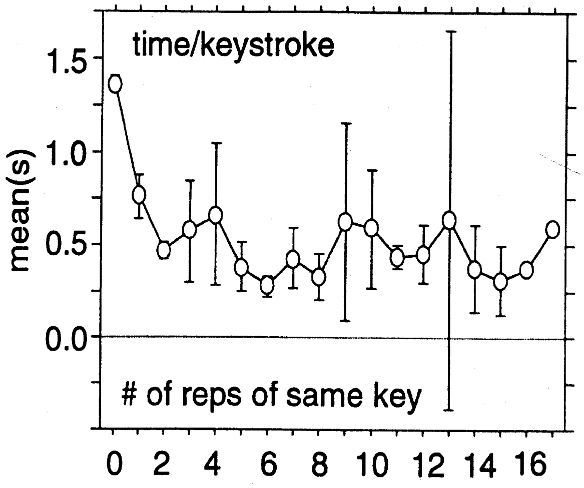
(Subject has typed the destination name)

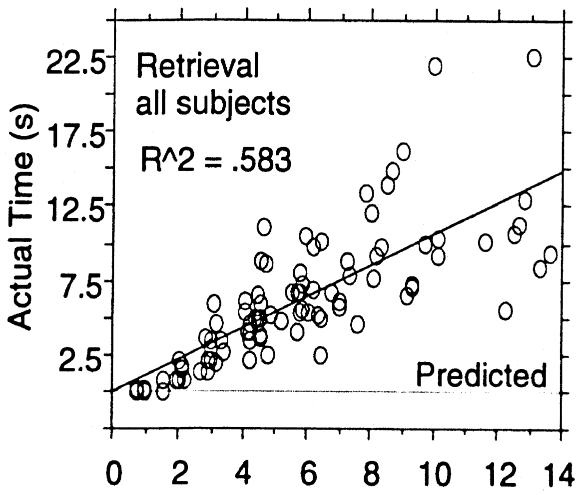


(Subject has typed the coordinates)

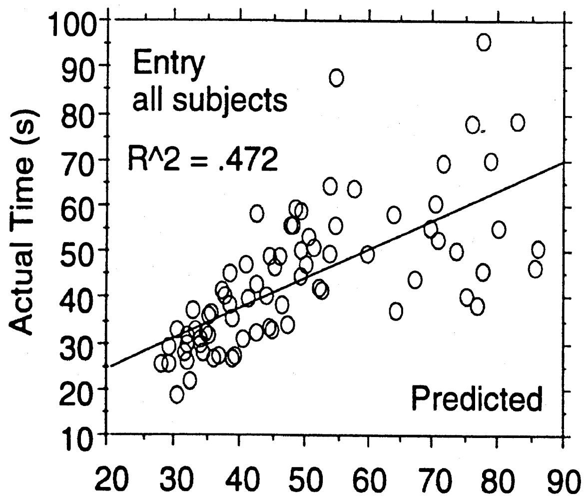
3 RESULTS







Actual Time = .039 + 1.054 *Predicted



Actual Time = 12.288 + .644 * Predicted

4 CONCLUSIONS - Keystroke Model Adjustments to Consider

Step 1: Revised Values for K (s)

Key Category	---Repetition--- 1st	2nd	>2nd
Cursor	1.71	0.69	0.47
Enter	1.55		
Letters	1.55	0.99	
Numbers	1.15	0.47	
Shift	1.46		
Space	0.60		

Step 2: Age Multiplier

Young	1.0
Middle	1.4
Older	2.2

Most important adjustment

Step 3: Lighting Multiplier

Dusk	0.94
Night	1.06

Step 4: Adjust Mental time

M = 2.22 s

PREFACE

This report is one of a series supported by the Road Commission of Oakland County, Michigan and the Federal Highway Administration, as part of the FAST-TRAC (Faster and Safer Travel through Traffic Routing and Advanced Controls) project. (See Underwood, 1994; Eby, Streff, Wallace, Kostyniuk, Hopp, and Underwood, 1996; Taylor and Wu, 1995; Kostyniuk, and Eby, 1996 for related research.) This operational field test combines the SCATS (Sydney Coordinated Automatic Traffic Control System) equipment and software, the Autoscope video detection system, and the Ali-Scout (Autofahrer Leit und Information System Scout) dynamic route guidance system. The goals of this effort are to improve traffic flow and reduce traffic accidents in Oakland County and the surrounding area.

Ali-Scout is a second-generation product developed by Siemens which provides real time, turn-by-turn guidance to drivers who have units installed in their vehicles. Ali-Scout vehicles communicate with infrared roadside beacons, which send travel times to the traffic control center and receive sequential routing instructions in return.

If navigation products are to be produced, they must be safe and easy to use. Driver navigation-related tasks include (1) calibration (of the compass and distance sensors) and setting (of the voice levels and screen colors), (2) telling the system where the driver wants to go (destination designation), and (3) following the guidance instructions. The second and third tasks are more important as most calibration and setting tasks requiring driver intervention are not performed while the vehicle is in motion. The human factors work carried out in the FAST-TRAC project is described in five reports, two of which relate to destination designation. A full description of an experiment involving destination designation using the Ali-Scout and subsequent analysis is provided in Steinfeld, Manes, Green, and Hunter (1996). This report describes models that predict destination entry and retrieval times through the use of individual keystroke times. Of interest is how long it takes drivers to determine the coordinates for a new destination, to enter coordinates into the navigation computer, and to retrieve previously entered destinations, and the time and errors for each. In addition, these two reports address a larger, more fundamental scientific issue—whether a touchscreen simulation of the real product is sufficient for usability assessments. The simulation takes much less time to construct and is easier to modify, facilitating iterative design.

Readers should note that the working title of this report was slightly different (Prediction of Destination Entry and Retrieval Times Using GOMS). While the acronym GOMS is commonly used in the human-computer interaction literature, the acronym and its meaning is unfamiliar to those involved with automotive human factors. Accordingly, a more readily understood term that also appears in the literature (keystroke-level model) was substituted.

Research relating to following route guidance is covered in three reports: one concerning equipment used in the evaluation (Katz, Green, and Fleming, 1995), one concerning driving performance and subjective ratings (Katz, Fleming, Green, Hunter, and Damouth, 1996), and a third concerning driver eye glances (Manes, Green, and Hunter, 1997, in progress).

This series of reports provides a comprehensive examination of driver-related design issues and should be useful in designing and evaluating safe and easy-to-use navigation products.

We thank Amitaabh Malhotra, Patrick Wei, and Marie Williams, all formerly of UMTRI, for programming the Ali-Scout simulation, and Aaron Steinfeld for significant contributions to the design of this experiment. We also would like to thank Sara Naylor for testing some of the pilot subjects.

Finally, the authors thank Cale Hodder of Toyota for encouraging the authors to include Americanized A3 reports (the two-page summary prior to the preface) in our technical reports.

TABLE OF CONTENTS

INTRODUCTION

Why this topic is of interest

There is considerable worldwide interest in developing safe and easy to use navigation systems for motor vehicles. Navigation systems typically include (1) a display capable of showing maps of the route and turn-by-turn guidance (usually in the form of arrows), and (2) computer-generated speech to provide directions. Such products are quite popular in Japan (Treece, 1996). Efforts to market navigation systems on a similar scale in the U.S. and Europe are just beginning.

Navigation systems can reduce wasted travel, saving drivers time and fuel, and provide for operational efficiency by optimizing use of the road network. By decreasing driving under uncertain conditions and eliminating the use of paper maps while driving, accidents may be reduced. Finally, navigation systems will offer comfort and convenience to drivers. However, such positive outcomes are predicated upon the assumption that navigation systems are safe and easy to use.

There are two primary driver tasks in using navigation systems: (1) entering and retrieving destinations, and (2) following the directions given by these systems (route guidance). Secondary tasks include setting and calibrating the system. Route following deserves the most attention because that task occurs while the vehicle is in motion. Route following is covered in other reports in this project (Katz, Green, and Fleming, 1995; Katz, Fleming, Green, Hunter, and Damouth, 1996).

However, destination designation also must be considered. If drivers cannot readily identify destinations to the navigation computer, there will be no guidance. Generally, destination designation is assumed to be performed while the vehicle is stopped or parked. However, in many circumstances, such as driving on an expressway, stopping may be difficult, so destination designation while in motion may be less risky. There is great concern as to what a driver can do while in motion (Zwahlen and DeBald, 1986; Zwahlen, Adams, and DeBald, 1988).

Previous research

Several studies in the literature have examined the entry of location names, street addresses, and coordinates, a focus of this experiment. For a detailed review of that topic, see the previous report in this series (Steinfeld, Manes, Green, and Hunter, 1996). Of those studies, several have attempted to predict performance times using a keystroke-level implementation of the GOMS (goals, operators, methods, and selection rules) model developed by Stuart Card and his colleagues at Xerox (Card, Moran, and Newell, 1983).

Summary of GOMS

GOMS was originally developed as a method for predicting the time to complete <u>routine</u> cognitive operations using a computer system. The model draws upon knowledge from psychology, industrial engineering, and computer science. The model is not intended to provide exact predictions of task times but reasonable approximations. The model assumes that the user is reasonably familiar with what they need to do (hence the routine term) and that the primary constraints are related to thinking (cognitive activities) as opposed to physical limitations (aerobic capacity, muscle strength, etc.).

The application of the GOMS approach can best be described by example. Suppose, for example, the task was to set the clock on a VCR. On a typical system, this would involve entering "clock" mode, setting the values for several fields (such as day, hours, and minutes), and finally exiting "clock" mode. An analyst might identify one top-level goal ("Set clock") and three subgoals ("Change mode," "Set field value," and "Advance to next field"). For each low-level goal (goals at lower points in the hierarchy), a method must be specified using basic operators (e.g., "Push up arrow button" or "Decide if value is correct"). Finally, selection rules are invoked any time more than one path can be taken. One could enter the hours either with the number pad or using the up and down arrow buttons.

Once the steps have been identified, the task times can be predicted using data from either the Model Human Processor or the Keystroke-Level Model. The Model Human Processor is a computer system representation of the human thought process consisting of three memory systems and four processors. These components are organized into three subsystems (perceptual, cognitive, motor). (See Figure 1.) Memories have three parameters: storage capacity (μ), storage code (κ), and decay time (δ). Processors have one parameter, cycle time (τ). For each parameter, three values are given: middleman, slowman, and fastman. Middleman is the most typical value. Slowman and fastman represent reasonable minima and maxima. The model also includes eye fixation times. The choice of the value depends upon the prediction needed—best case, worst case, or typical.

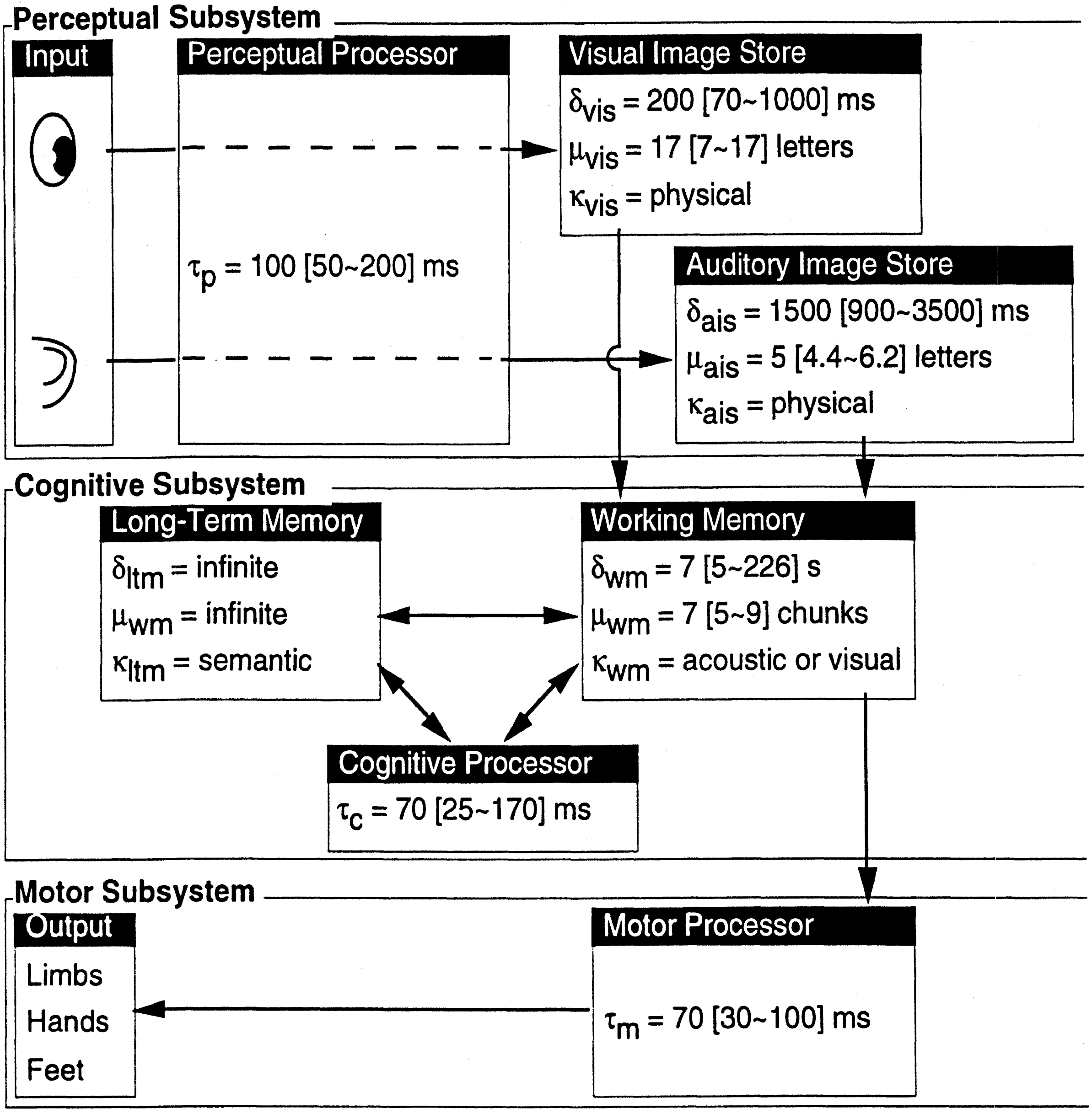


Figure 1. Overview of the Model Human Processor.

A simplified GOMS derivative is the Keystroke-Level Model (Card, Moran, and Newell, 1980). The Keystroke-Level Model is higher level than the Model Human Processor in that each element in the Keystroke-Level Model can be decomposed into several elements in the Model Human Processor. Accordingly, less time is required to estimate times using the Keystroke-Level Model. However, the estimates are less accurate because some details are ignored.

Model parameters are shown in Table 1. Two parameters are of particular interest: (1) K, the time to type a keystroke (the interkeystroke interval) and (2) M, the time to mentally prepare to perform some action such as typing a digit sequence or moving a mouse to a location. Card, Moran and Newell (1980) provide a detailed set of rules identifying where M parameters should be included in time estimates. (See also Kieras, 1988.)

Table 1. Keystroke-Level Model parameters.

Parameter	Symbol	Comment	Time (s)
Pointing	P	Point with a mouse to a target on a display	1.1
Homing	H	Home hand(s) to keyboard or to device	0.4
Draw	D	Draw N straight lines of length L cm	.9N + .15L
Mental	M	Mentally prepare	1.35
System Response	R	System specific time, empirically determined	t
Keystroke	K	Best typist (135 wpm)	0.08
		Good typist (90 wpm)	0.12
		Average skilled typist (55 wpm)	0.20
		Average nonsecretary typist (40 wpm)	0.28
		Typing random letters	0.50
		Typing complex codes	0.75
		Worst typist (unfamiliar with keyboard)	1.20

Olson and Nilsen (1987 and 1988) carried out and analysis of what users do in working with spreadsheets. From their data, M (the mental or thinking time) is estimated to be 1.62 seconds and K (time per keystroke) to be 0.36 seconds. In addition, they also introduce a new parameter, S (equal to 2.29 seconds), which involves scanning across a row to find an item in a matrix.

Previous studies of data entry modeled using GOMS

Detweiler (1990) examined five different methods to enter characters using a 12-key telephone pad. As shown in Table 2, the rank orders of the actual entry times and keystroke-level predictions were identical with one exception. Notice that all the differences between the actual and predicted times for each method are under 10 percent and three of five are under five percent. Keystroke-level models were more likely to underpredict than overpredict.

Table 2. Entry times and errors for various methods from Detweiler (1990).

Method	% Errors	Entry Time (s)	GOMS (s)	Difference (%)
Repeat key	6.7	12.38	11.96	3.4
Modified modal	17.1	12.50	13.72	9.8
Top row	8.0	13.50	13.50	0.0
Same row	10.5	14.18	13.78	2.8
Modal position	13.0	14.81	14.58	1.6

In an experiment even closer to the point, Paelke (1993) examined four different interfaces for destination entry spanning the range of options available in contemporary navigation systems. Table 3 shows data from her efforts. While the percentage differences are larger, that, in part, was because of larger sources of variance in entry times (time sharing, slow response of the simulated interface, etc.).

Table 3. Entry times and errors for various methods from Paelke (1993).

Method	Similar to	Entry Time (s)	GOMS (s)	Difference (%)
Doublepress	TravTek	75.6	63.1	16.5
Qwerty	ADVANCE	44.1	54.2	22.9
Phonepad		42.9	49.8	16.1
Scrolling List	Zexel/PathMaster	55.1	82.5	81.0

Nonetheless, in both studies, the rank order of keystroke-level predictions agrees with the actual data, suggesting keystroke-level predictions are useful for choosing among alternative interface designs. This has tremendous importance for interface design and evaluation because an experimental evaluation of alternative interfaces might take several months to complete, while keystroke-level calculations could be completed in a few days to a week. About 40 percent of the time spent testing an interface is spent planning the experiment and instrumenting the interface or developing a simulation.

Research issues explored

The research described here provides an example of the application of keystroke-level models to two tasks involving a fairly complex interface. While the previous Ali-Scout report focused on how various factors (subject age and sex, lighting conditions, and stimulus) affected overall task times, this report deals with entry and retrieval times at a more fine-grained level. Specifically, the following two issues were addressed:

1. What are typical keystroke times when using an interface with an unfamiliar keyboard and small buttons?
2. How successfully can keystroke-level models predict overall times for tasks of varying complexity?

The following two sections provide overviews of the relevant features of the previous Ali-Scout study, namely the basic experimental protocol and some of the key results. Next, the results of the in-depth keystroke analysis are discussed. The final two sections describe the development of keystroke-level models for two tasks and an analysis of how reliably the models predicted the actual task times from the experiment.

TEST PLAN

This section provides a summary of the test protocol. Additional details appear in the first report in this series (Steinfeld, Manes, Green, and Hunter, 1996).

Test participants

There were 36 subjects in the experiment—12 young (18 to 30), 12 middle-aged (40 to 55), 12 older (over 65)—representing the population extremes and the most likely buyers. Within each age group there were an equal number of men and women. All subjects were licensed drivers, driving between 1,000 and 40,000 miles per year (mean of 13,000). All but three subjects had at least some college experience. Subjects' visual acuity ranged from 20/13 to 20/40. Only one subject had previous experience with a navigation system, but most had used a touchscreen. Computer use was moderate on average. Subjects were moderately comfortable typing. Subjects were paid $40 for completing the experiment.

Test materials and equipment

Ali-Scout interface

This experiment incorporated a real Siemens Ali-Scout Display Unit as well as a touchscreen simulation created in SuperCard on a Macintosh computer. Figure 2 shows frames grabbed from video recordings of device use. The Ali-Scout interface consists of the following four elements: (1) an LCD guidance screen on the left of the unit face; (2) a text window for destination names, coordinates, and entry information; (3) front panel selection keys; and (4) a fold-out alphanumeric keypad.

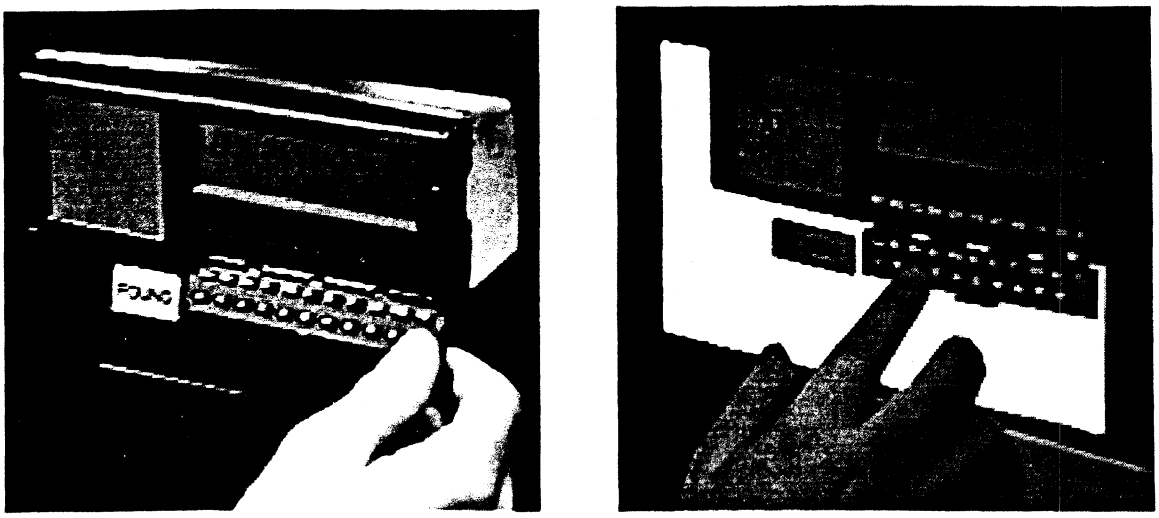


Figure 2. The real display unit (left) and the simulated display unit (right).

Figures 3 and 4, actually taken from the simulation, show the Display Unit closed and open. These figures are full size on an 8.5 x 11 inch page. The FOUND button was not part of the device but was added to assist in timing use performance. The size and appearance of all elements of the simulated display were identical to the real interface except that there was no tactile feedback when a key was pressed. A tone was presented instead. Notice the small key size.

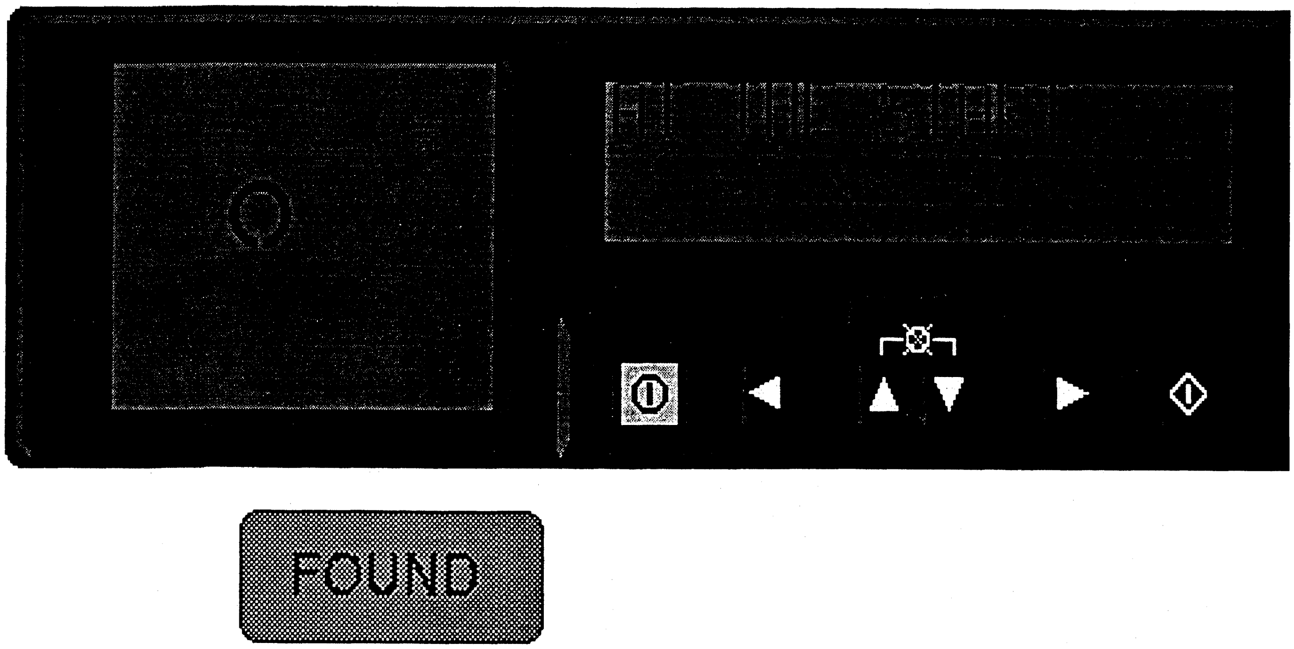


Figure 3. The simulated display unit with the door up.

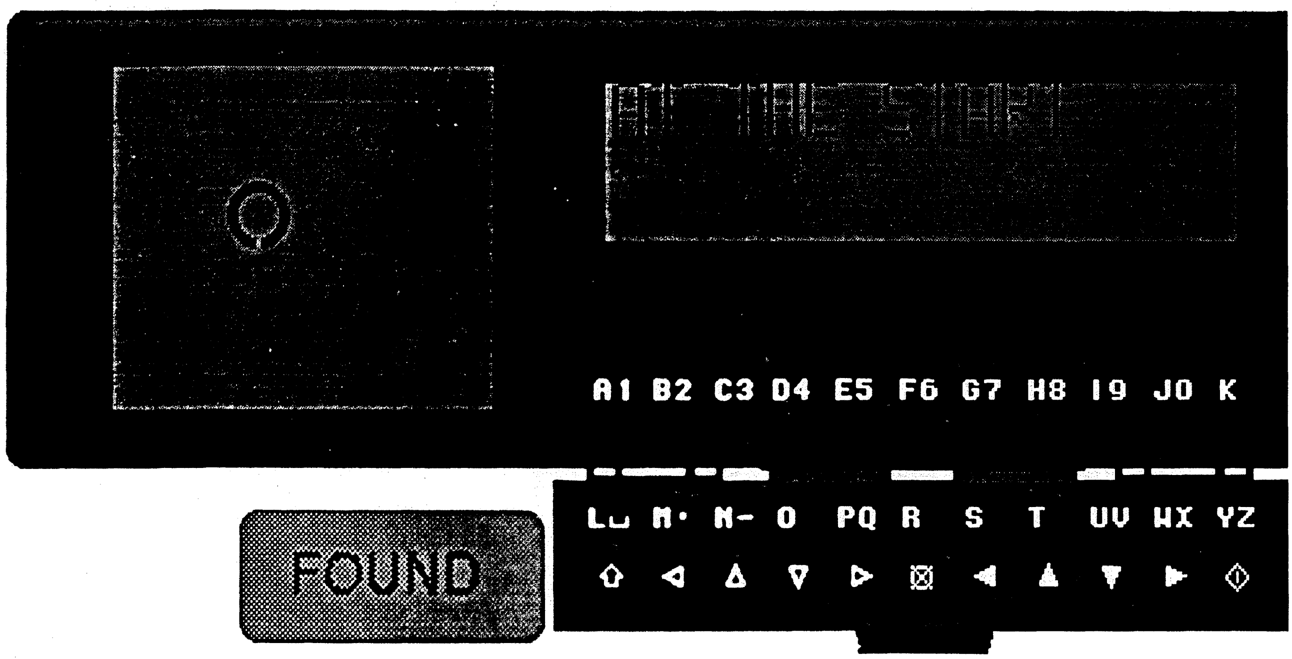


Figure 4. The simulated display unit with the door down.

The real Display Unit was mounted on a flexible stalk with the display face positioned 3.5 inches (8.9 centimeters) in front of the touchscreen when in use (see Figure 5).

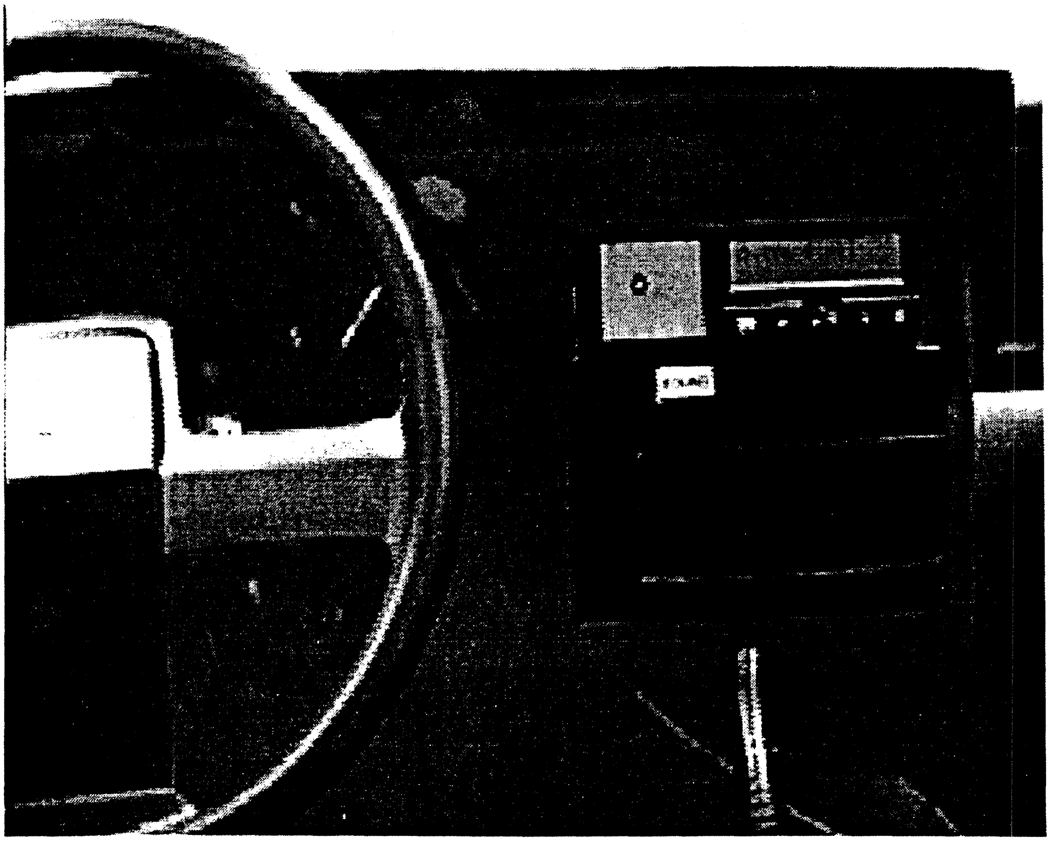


Figure 5. Location of the display.

Two tasks were examined: (1) retrieving previously entered (stored) destinations and (2) entering new destinations. To retrieve a destination, subjects could use one of three strategies: (1) type in the name of the destination, which appeared when the characters entered uniquely matched the beginning of that name (type method); (2) scroll through the list of names until the desired destination appears (scroll method); or (3) type in one or more characters and then scroll the rest of the way (hybrid method).

Table 5 shows the keystrokes necessary to obtain SEARS as a destination using the type method and hybrid method. Assuming the subject had not memorized the data base (there were 21 locations in the main list), the minimum character strokes required using either method would be four. The first down arrow is required to enter the scrolling mode.

Table 5. Example of the type and hybrid methods for destination retrieval.

Location List	Type Method		Hybrid Method	
AT THE START	Entered	Displayed	Entered	Displayed
...	S	S......	S	S......
SAKURA BANK	E	SE.....	∇ (scroll down)	SAKURA BANK
SEAFOOD BAY	A	SEA....	∇	SEAFOOD BAY
SEARS	R	SEARS	∇	SEARS
VANDENBURG SCH				
...				

To enter a destination not in the data base, the subject first entered the name (up to 14 characters including spaces). Keying was somewhat confusing as many of the keys had two characters on them. The left character was shown in white, the right in yellow. To type the right character, the subject first pressed the shift key (which was located at the bottom-left key of the keyboard and labeled with a yellow up arrow) and then the key of interest. (See Figure 4.)

Next, the subject keyed in the longitude and latitude. This involved advancing the cursor to each of the two coordinate fields using either the diamond key or the right cursor key and typing the appropriate numbers. Finally, the subject pressed the diamond key to save the destination. The longitude and latitude were obtained either from a map or from lists of street address ranges and coordinates. Additional details of the entry process are provided later.

The original project plan called for evaluating both real and simulated Display Units under simulated dusk and night conditions. However, pilot tests showed no differences due to illumination for the simulated unit, so only the simulated dusk condition was explored in the main experiment.

Driving simulator

The data collection portion of the experiment was conducted in the UMTRI Driver Interface Research Simulator. The automobile simulator consisted of an A-to-B pillar mockup of a 1985 Chrysler Laser, a retroreflective wall, and a variety of computer and video components. Subjects never drove the simulator. During experimental trials an image simulating the view out the windshield of a car parked in the right shoulder of a two-lane road was presented. For additional details of the simulator, see the initial report, Green and Olson (1997), Olson and Green (1997), or MacAdam, Reed, and Green (1993).

Test activities and their sequence

An overview of the protocol is shown in Table 6. As noted in the table, the primary test tasks were to retrieve and enter three sets of five destinations. Locations retrieved were ordered so that the minimum number of keystrokes (averaged across groups by trial) was just over three (see Tables 7 and 8). Likewise, destinations entered were ordered to roughly equalize the total number of keystrokes and shifts across orders and across entry trials to facilitate looking at differences due to those factors. (See Table 9.)

Table 6. Summary of the experiment.

Activity	Name	Description
1	Introduction	The subject was told the purpose of the experiment, and then completed the biographical and consent forms.
2	Videotape	The subject watched an instructional video on entering and retrieving destinations.
3	Practice	The subject retrieved five locations, then entered five locations.
4	Test-use of manual	The subject looked up three destinations in the manual (point of interest name, intersection of 2 roads, street address).
5	Simulator introduction	The subject practiced using the touchscreen.
6	Test-keypad use	The subject completed five entry then five retrieval tasks (three times: real interface at dusk, real at night, simulated at dusk).
7	Posttest	The subject's eyesight was checked, completed a questionnaire, was paid, and finger anthropometry was recorded.

Table 7. Retrieval lists for each stimulus set.

A	B	C	Dummy
SAKURA BANK	SEAFOOD BAY	BILL KNAPPS	MONTGMRY WARD
BIR ICE ARENA	PRINT GALLERY	PRIMOS PIZZA	ROYAL OAK DELI
MONTERREY REST	MAJESTIC CAFE	WOODSIDE HOSP	SEARS
MOBIL	VANDENBURG SCH	BIR THEATER	BIR ART GALLRY
BIG BOY	BIR LIBRARY	MONGOLIAN BBQ	PALACE OF AH

Table 8. Minimum number of keystrokes for retrieval.

Stimulus Set			
A	B	C	Dummy
2	3	3	5
5	4	3	1
5	2	1	4
3	2	5	4
2	5	4	2
3.4	3.2	3.2	3.2

Table 9. Destination lists for each entry stimulus set.

Stimulus Set	Name	Total Keystrokes	Total Shifts
A	NICKS PLACE	12	1
	Q GAS	7	2
	HELENS KITCHEN	15	1
	YAW GALLERY	12	1
	GOODYEAR	8	0
B	FARMER JACK	12	1
	TACO LOCO	10	1
	FIRST OF AM	13	2
	JACOBSONS	9	0
	CHEVRON	8	1
C	LARK REST	10	1
	UNICORN GRILL	14	1
	KROGERS	7	0
	QWIK STOP	11	2
	TUFFY AUTO	11	1
Dummy	LICHT PARK	11	1
	NORDSTROM	9	0
	DISCAFE	7	0
	OAKLAND MALL	13	1
	OLIVE GARDEN	14	2

Each set of five was entered under different conditions (real interface at dusk or at night, simulated interface at dusk) in a counterbalanced order. Lighting conditions were simulated by adjusting the interior lighting.

SUMMARY OF THE RESULTS FROM THE PREVIOUS REPORT

Keypress times for the simulated interface were automatically saved by the controlling software. Keypress times for the real interface were obtained by playing back a videotape of user actions at reduced speed and by using a custom logging program to enter keystrokes.

Task times varied quite widely with the task. Figures 9 and 10 show the histograms of times for retrieval for both tasks. There were significant differences between destinations (presumably due to the number of keypresses required), an issue that will be explored later. (See Table 16.) Also prominent were practice effects (primarily between the first and second block), and significant effects due to age, sex, and subject within age-sex categories. Lighting was also a significant factor (dusk versus night), but only for older subjects performing the entry task.

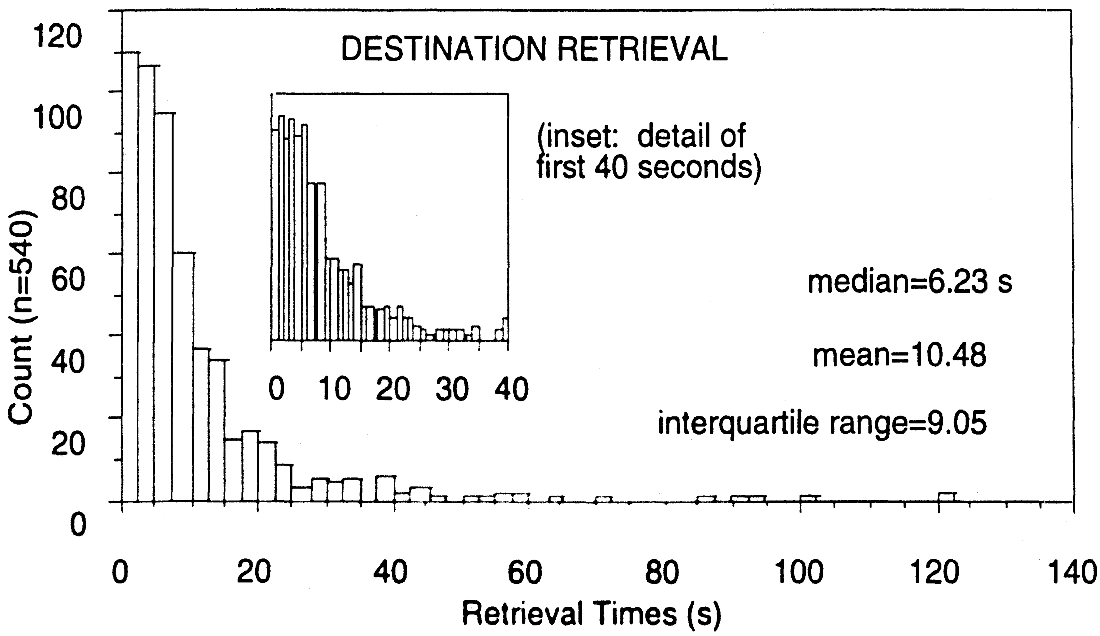


Figure 9. Histograms of retrieval times.

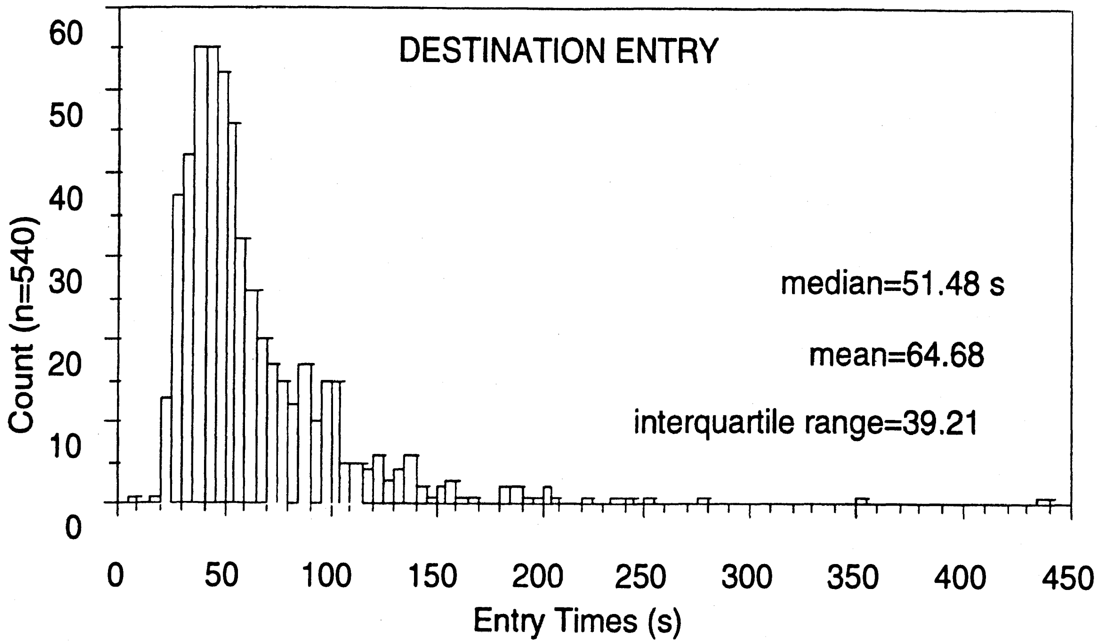


Figure 10. Histogram of entry times.

Table 16. Median, maximum, and mean times for each of the 30 cards.

Retrieval Times (in seconds)				Entry Times (in seconds)			
Place Name	Median	Max	Mean	Place Name	Median	Max	Mean
SAKURA BANK	3.47	101.88	8.80	NICKS PLACE	60.16	209.95	77.00
BIR ICE ARENA	12.00	121.75	19.00	Q GAS	47.67	275.88	69.55
MONTERREY REST	9.58	120.58	16.34	HELENS KITCHEN	55.25	203.22	72.85
MOBIL	3.98	55.40	8.68	YAW GALLERY	54.19	252.55	67.82
BIG BOY	4.57	51.83	7.47	GOODYEAR	39.45	135.73	51.05
SEAFOOD BAY	8.27	94.90	17.10	FARMER JACK	67.55	351.30	84.31
PRINT GALLERY	6.82	16.05	7.52	TACO LOCO	52.50	151.70	57.98
MAJESTIC CAFE	2.37	21.30	3.73	FIRST OF AM	54.98	159.20	64.55
VANDENBURG SCH	3.56	52.97	9.68	JACOBSONS	48.28	185.70	60.30
BIR LIBRARY	9.97	58.13	14.05	CHEVRON	44.00	100.63	46.96
BILL KNAPPS	4.79	62.60	8.70	LARK REST	61.74	243.73	72.03
PRIMOS PIZZA	6.15	57.63	10.93	UNICORN GRILL	55.59	238.53	74.17
WOODSIDE HOSP	0.38	19.35	1.83	KROGERS	43.96	101.52	47.28
BIR THEATER	9.83	30.77	12.70	QWIK STOP	54.43	436.45	73.18
MONGOLIAN BBQ	7.27	39.77	10.69	TUFFY AUTO	45.67	131.72	51.12
Overall	6.23	121.75	10.48	Overall	51.48	436.45	64.68

Note: Retrieval always begins with the place name AT THE START (the first alphabetic entry) shown.

Also important was the relationship between the retrieval and entry times for real systems (both at night and at dusk) and analogous times for simulated systems. In general, actual task times were 57 to 86 percent of the time using the simulated interface, depending upon the lighting condition. Some of this difference may be due to the slowness of the interface in responding to input, something that should not be an issue with current, much faster, computers. However, the lack of tactile feedback provided by the screen is still an issue.

The error data were also addressed in some detail, but the number of errors was relatively small, so few differences were found. Left unanswered in the previous report was how the details of the task structure affected keying time.

1. Were the times for different types of keystrokes the same?
2. How did individual differences affect the time for each type of keystroke?
3. How did the simulated keystrokes differ from real keystrokes
4. How well do the GOMS predictions agree with the times measured

   for the real device (dusk, night),
   for the simulated device, and
   for each task?

This report focuses on issues 1 and 4, the differences between keystroke types and the ability of keystroke-level models to predict actual performance.

ESTIMATING MODEL PARAMETER TIMES USING THE EXPERIMENTAL DATA

One of the major goals of this project was to estimate the operator times using actual data from the experiment. Although keystroke-level models can provide predictions of task times based on the design of a product so testing is avoided, initial keystroke-level estimates for the Ali-Scout were very rough. This is partially because standard keystroke operator times in the literature are based on the QWERTY keyboard which has very different characteristics than the keypad of the Ali-Scout, so it was not apparent which estimate of keying time should be used. Also, the nature of the tasks performed using the Ali-Scout may be sufficiently different from those used in the literature that the standard mental operator times do not apply. Since it was unclear which keystroke-level parameter estimates (or others tailored from the data) were most accurate, an effort to provide model calibration data for this context was appropriate. Again, the ultimate goal was to use some form of keystroke-level model as a basis for comparing alternative driver interfaces during interface development.

The basic approach here was to examine each main effect (such as keying numbers versus letters) and interaction in a stepwise fashion to reveal patterns in the data, especially for categorical effects such as age. To ensure that keystroke times met the assumptions of the GOMS model approach, only perfectly executed responses for the real (rather than simulated) Ali-Scout interface were examined. Furthermore, four exclusion criteria were applied to the data set for the real device.

1. No extra keystrokes - Sequences in which extra keystrokes appeared were trials for which subject uncertainty was likely. Uncertainty violates the routine-cognitive-method assumption of GOMS. Further, accounting for the corrections would have required developing a specific GOMS model for each trial with extra keystrokes, an excessively time consuming task.

2. No first keystroke - Because of the method of timing, the first keystroke time also includes the time for a mental operator. There was no way to determine exactly when the subject began to plan the keying sequence and when planning ended. However, when the first keystroke was completed was known.

3. No first numbers - The time for the first number in a sequence has embedded in it the time for a mental operator since the subject plans a number sequence just before typing the string.

4. Young subjects only - Young subjects (1/3 of the sample) were least likely to pause while keying to think about what to do next, and accordingly for them, keying was close to being a routine cognitive activity. For middle-aged and older subjects, pauses did not occur in the same places for all subjects and dropped out with practice, making it difficult to determine where mental operators should be included in the time estimates.

The need for these assumptions is apparent in the original data set, where a moderate number of operator times were in excess of 10 seconds, times well beyond keystroke-level model estimates.

Using those criteria, 2,181 data points remained from 12,107 in the original data set. These remaining trials do not represent optimal performance, particularly for retrieval tasks where several methods could be employed. The mean time per operator was 1.24 seconds with a standard deviation of .93 seconds. Figure 11 shows the distribution of times and Figure 13 shows the distribution of log (time). Clearly the distribution is log normal.

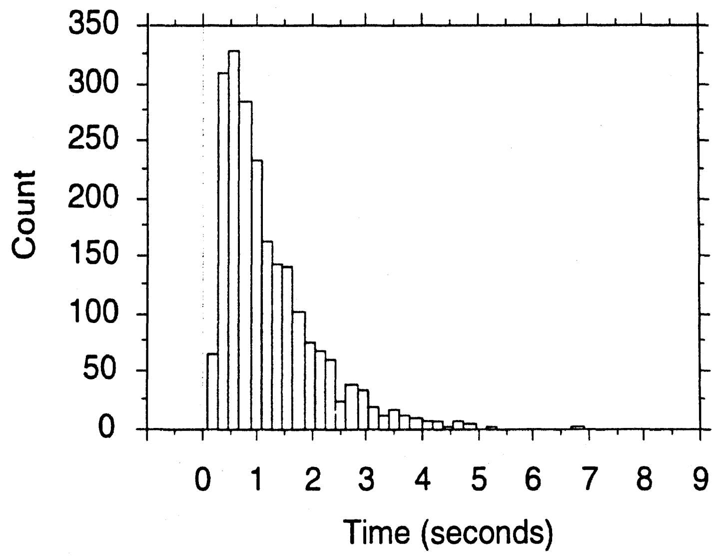


Figure 11. Distribution of times for young subjects with no extra or initial keystrokes.

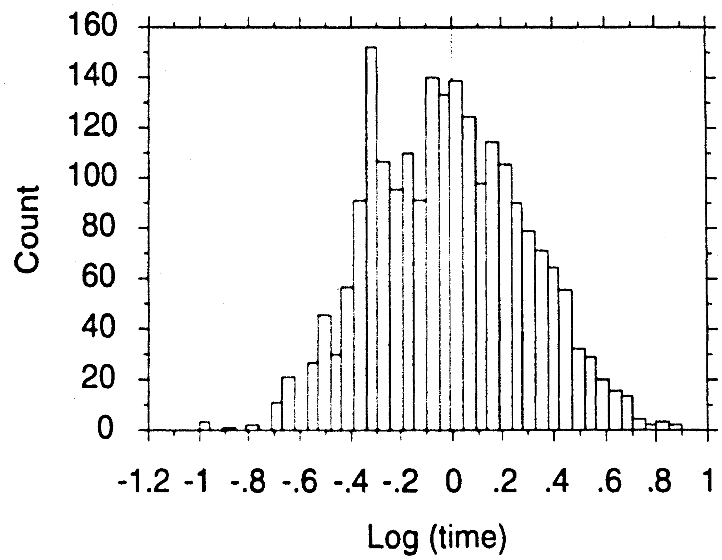


Figure 12. Distribution of log (time) for young subjects with no extra or initial keystrokes for real interfaces.

Effect of character typed

There is considerable data in the literature on entry of sequences using a single finger (e.g., Verwey, 1996) and typing (e.g., Evey, 1980, Cooper, 1983). However, the keying process in this context was only partially automated, and context-specific data was therefore desired.

Table 17 shows a summary of the keystroke times by key. Of the characters typed, only 10 of them had fewer than 20 entries, so the time estimates should be fairly stable. Table 18 shows those same data sorted by mean time. Time for "other" keys (cancel, etc.) are not shown as those keys were associated only with extra keystroke sequences.

Table 17. Time statistics for each key, sorted by key name.

Key Name	Mean	Std. Dev.	Std. Error	Count	Minimum	Maximum
0	1.039	.513	.058	78	.300	2.730
1	1.024	.614	.120	26	.470	3.030
2	.780	.396	.051	61	.270	2.200
3	1.230	.888	.090	98	.200	5.300
4	1.155	.877	.127	48	.230	3.930
5	1.218	.613	.084	53	.330	3.000
6	1.188	.811	.127	41	.400	4.170
7	1.032	.671	.101	44	.430	3.570
8	1.438	.925	.164	32	.470	4.270
9	1.150	.928	.124	56	.470	4.800
A	.943	.592	.063	88	.300	4.470
B	1.627	.690	.208	11	.800	2.700
C	1.134	.583	.088	44	.430	2.900
D	.990	.416	.208	4	.430	1.430
E	1.300	.653	.083	62	.430	3.500
F	1.498	1.236	.258	23	.270	4.730
G	2.188	1.066	.175	37	.470	5.530
H	2.274	1.131	.400	8	1.100	4.370
I	1.842	1.153	.128	81	.470	5.870
J	2.727	1.027	.388	7	1.400	4.630
K	1.585	.909	.163	31	.330	3.700
L	1.616	1.097	.152	52	.170	4.700
M	1.344	.639	.160	16	.730	3.270
N	1.403	.925	.136	46	.170	4.100
O	1.557	1.111	.124	80	.270	7.370
P	1.934	1.044	.290	13	.400	3.500
Q	.834	.302	.084	13	.530	1.570
R	1.677	.883	.088	100	.400	4.870
S	1.531	.994	.142	49	.430	4.830
T	1.121	.804	.134	36	.330	3.430
U	1.205	.755	.202	14	.430	2.970
V	.804	.116	.037	10	.670	1.030
W	2.215	.989	.211	22	1.070	5.130
Y	2.038	1.071	.260	17	.670	4.400
Right cursor	.917	.765	.052	216	.100	4.600
Scroll up	.470	.381	.071	29	.230	2.370
Scroll down	.666	.586	.053	124	.200	2.970
Enter	1.552	1.107	.085	170	.570	7.870
Shift	1.277	1.048	.085	153	.270	6.770
Space	.565	.222	.024	88	.230	1.400
Total	1.243	.932	.020	2181	.100	7.870

After very careful inspection of the data to identify similarities, and based on accepted conventions, the keystrokes were grouped into six categories—letters, numbers, cursor keys, enter, shift, and space. A preliminary examination of all letters versus letters on the first row (thought to be different because of differences in illumination level) showed no differences. In addition, there also did not appear to be a relationship between the frequency with which a character was typed and its mean time (r=-0.26 for the data set). Partitioning the data set (e.g., numbers only, letters only) led to even lower correlations between frequency and mean interkeystroke interval.

Table 18. Mean keying times for the real interface sorted by time.

Key Name	Time (seconds)	Key Name	Time (seconds)
Scroll up	0.470	Shift	1.277
Space	0.565	E	1.300
Scroll down	0.666	M	1.344
2	0.780	N	1.403
V	0.804	8	1.438
Q	0.834	F	1.498
Right cursor	0.917	S	1.531
A	0.943	Enter	1.552
D	0.990	O	1.557
1	1.024	K	1.585
7	1.032	L	1.616
0	1.039	B	1.627
T	1.121	R	1.677
C	1.134	I	1.842
9	1.150	P	1.934
4	1.155	Y	2.038
6	1.188	G	2.188
U	1.205	W	2.215
5	1.218	H	2.274
3	1.230	J	2.727

Figure 13 shows the mean times for each category of keys for each task. For convenience, those means also appear in Table 20. Note the high variability in the shift key times for retrieval in the figure, most likely because use of the shift key was associated with beginning a new sequence, and in some cases there may have been associated mental activity. Time for the space and cursor keys tended to be less than others because those keys are often struck several times in a row, and search to find the key was only required for the first keypress. These differences may also reflect frequency of use (and repeated use) as the shift and space keys were used quite often.

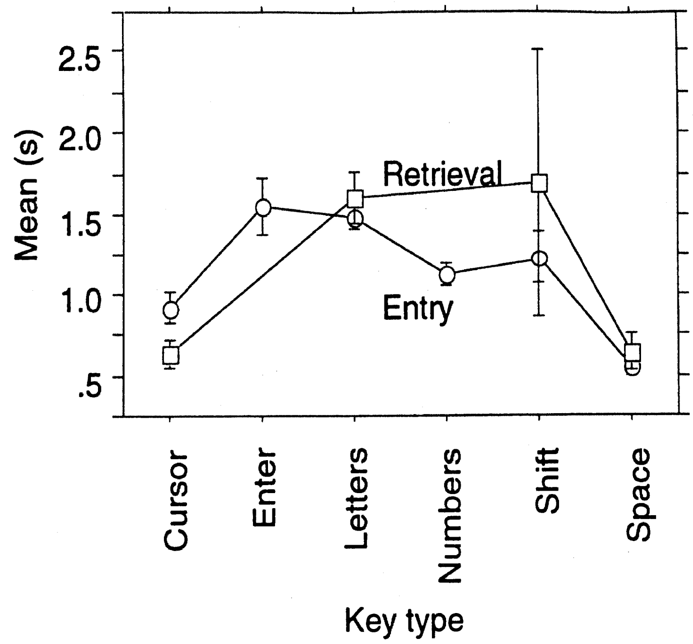


Figure 13. Times by key category for young subjects using the real Ali-Scout.

Table 20. Mean times for young subjects using the real Ali-Scout.

Key Category	Entry (s)	Retrieval (s)
Cursor	0.92	0.63
Enter	1.55	
Letters	1.49	1.61
Numbers	1.12	
Shift	1.23	1.69
Space	0.55	0.64

To put these values in perspective, the standard Keystroke-Level Model estimates for typing are 0.5 s/keystroke for random letters, 0.75 s/keystroke for complex codes, and 1.2 s/keystroke for a "worst" typist unfamiliar with a keyboard. The data obtained suggests times resembling that of a worst case typist, though there is considerable variation between key types, so using a single fixed keying time may not be optimal.

Effect of repetitions

As was mentioned previously, repetition was a critical factor. As depicted in Figure 14, the primary difference found was between the first trial and all others. The high degree of variance for repetition 13 was most likely due to outliers in the data set (only four data points existed for that number of repetitions). According to these data, the mean time for a keystroke was 1.37 seconds for the first keystroke, 0.76 seconds for the second keystroke, and 0.47 seconds for the others. The mean time for trial three and beyond is fairly close to the estimated time for typing random characters (0.5 seconds). Partitioning these data by key category (see Figure 15) shows that all of the repetitions greater than two were for the cursor key. The increase in time for the shift key probably was due to the complex planning required (since shift is part of a two-key sequence) and because of the low frequency of occurrence of that combination.

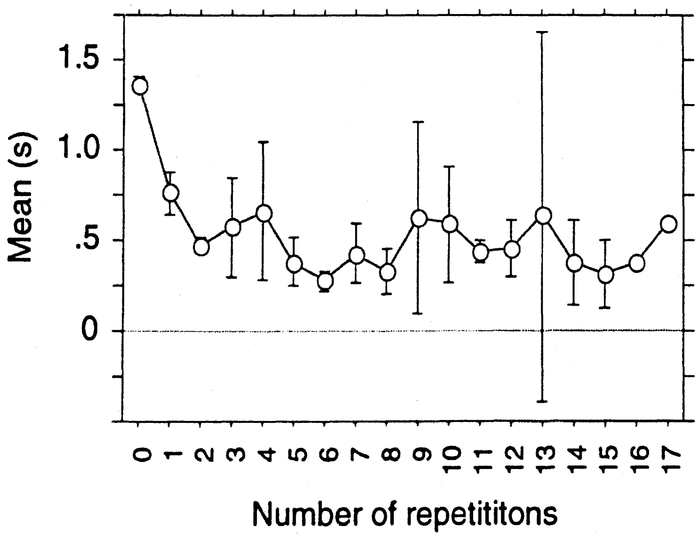


Figure 14. Effect of repetitions on the mean keying time (in seconds).

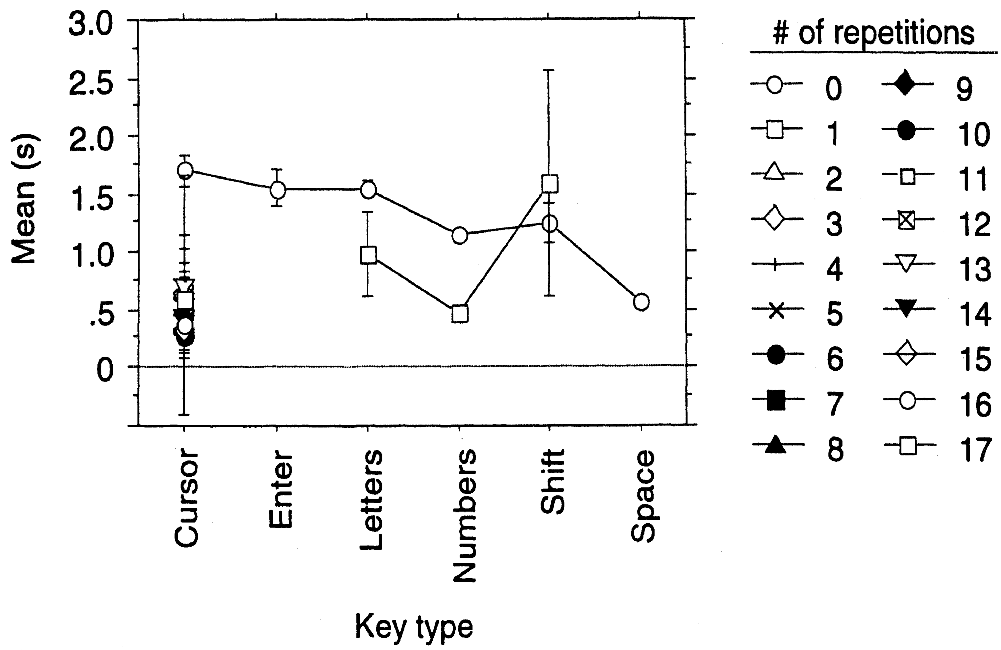


Figure 15. Mean keying time as a function of key type and number of repetitions.

Table 21 reveals that the relative decrease in the second key varies with the key category, though the means for more than two successive cursor keys and two successive numbers were both 0.47 seconds. This value probably represents a lower bound of keying performance for this device (on average for younger subjects).

Table 21. Mean times for successive keystrokes (in seconds).

	Keystroke		
Key Category	1	2	>2
Cursor	1.71	0.69	0.47
Letter	1.54	0.99	
Number	1.15	0.47	

Effect of dusk versus night

Also of interest is how the pooled dusk and night data described previously needed to be adjusted to account for dusk-night differences. Pooling was necessary to provide enough data for good estimates of the means. As shown in Figure 16, the major difference between dusk and night seems to be that enter, letters, and numbers took 10 to 15 percent longer to type at night than under dusk conditions. For the shift key, the reverse was true. Typing the shift key took slightly less time at night. This added time was needed to search for the key when illumination levels were low. Since the cursor key was often used several times in succession and the space was part of a two-key planned sequence, search was not required as often, so the effects of low illumination were much less. This suggests selective use of an adjustment factor for poor lighting conditions (decreasing the values in Table 20 by five to seven percent for enter, letters, and number to predict dusk performance; increasing them by the same amount to predict performance at night).

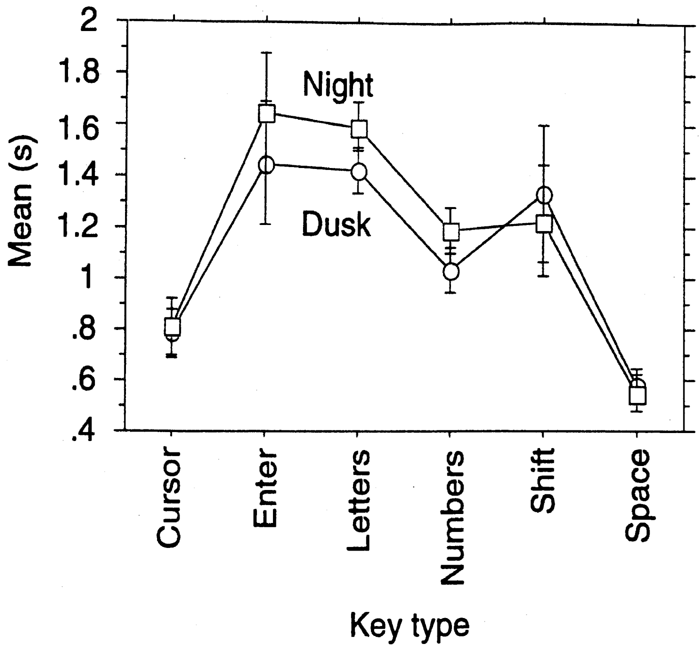


Figure 16. Keystroke times for real interfaces for young subjects—dusk versus night.

Effect of subject age

Age differences varied from one key type to another. (See Figure 17 and Table 23.) Although there were a limited number of good data points (trials without excess keypresses) for older drivers, note that times for enter and shift were considerably greater for that age group. This is probably because older subjects had not fully learned the entry process, so there was an added mental operator. Ignoring the shift and enter cases, times for older drivers were almost double (actually 1.93 times) those of young drivers.

If this were always the case, then adding in a standard M operator (1.35 seconds) would lead to an overestimate of the enter time (1.55 * 1.93 + 1.35 = 4.34 seconds) but would improve the accuracy of the shift times (1.28 * 1.93 + 1.35 = 3.82 seconds). Most likely, mental operations only occurred some of the time. For quick estimates, it is reasonable to assume that times for middle-aged drivers are 1.45 times those for young subjects when key types are equalized. Older driver keying times were roughly 2.19 times greater than young drivers. Again, in the experiment, most of the keys were letters and numbers. Shift and space, both of which had extreme values, were less common.

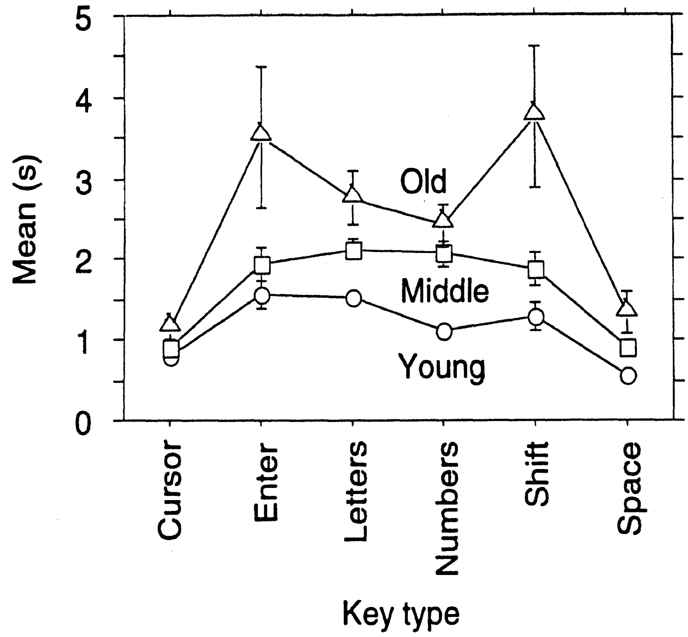


Figure 17. Age and key effects for real interfaces.

Table 23. Mean keying times as a function of subject age.

	Age		
Key Category	Young	Middle	Old
Cursor	0.80	0.90	1.13
Enter	1.55	1.94	3.50
Letters	1.51	2.13	2.75
Numbers	1.12	2.07	2.43
Shift	1.28	1.88	3.76
Space	0.57	0.89	1.33
Mean	1.13	1.64	2.48
Ratio to young		1.45	2.19

Figure 18 shows the relationship between age, key type, and task. There were no points for enter and numbers for retrieval since those keys were not required for that task. Cursor times were shorter for retrieval, perhaps because that key was used more frequently and often several times in a row. Older subjects appeared to take longer to type letters for retrieval than for entry, although this could be a result of insufficient data. Finally, middle-aged subjects took longer to press the shift key while retrieving destinations than while entering them, with times almost as long as the older subjects. However, note in Figure 19, the small overall difference between middle-aged and older subjects for destination retrieval.

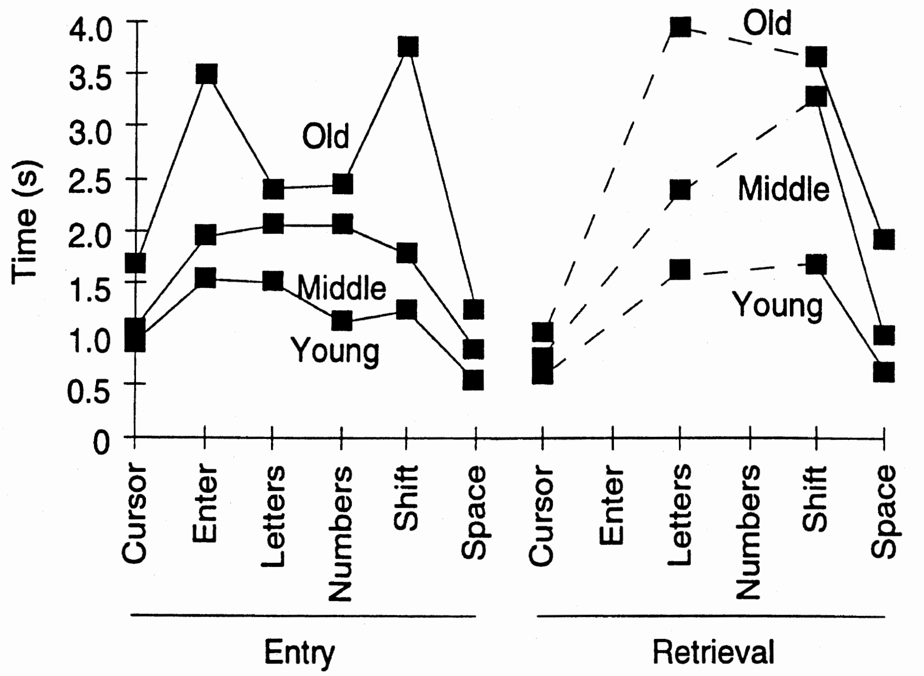


Figure 18. Keying times for real interfaces by task and key category.

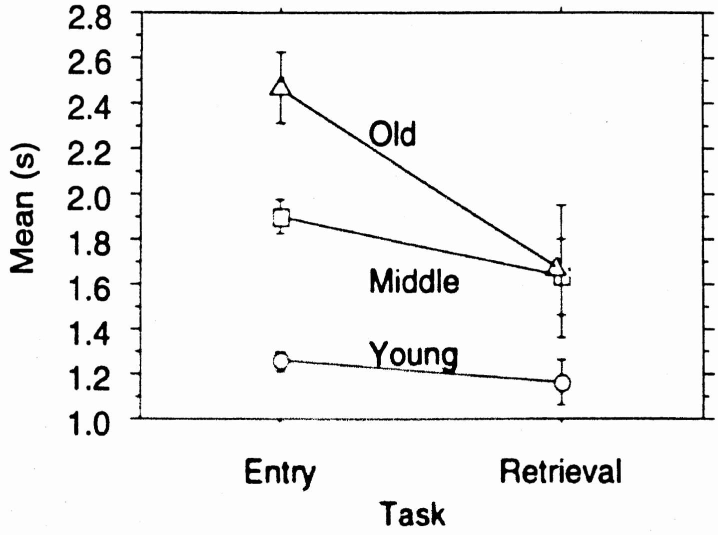


Figure 19. Interaction of task and age.

Estimation of mental time

There are several ways in which the time for a mental operation (M) can be estimated. In the entire data set, there were 154 instances where a digit sequence occurred but there were no extra keystrokes. This occurred for only five of the 10 digits. The mean time for that type of keystroke was 3.37 seconds. As shown in Figure 20, the primary difference is between 0 and other digits.

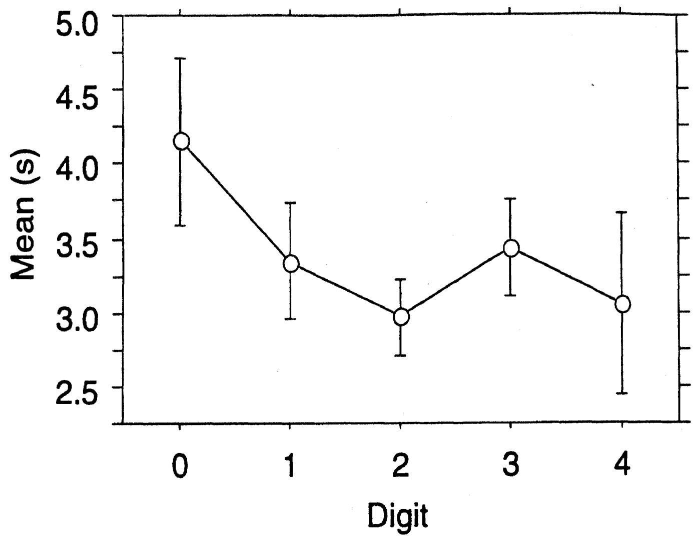


Figure 20. Times for initial digits by young subjects for real interfaces.

In a keystroke-level analysis, these sequences should be coded as one M (mental) operator followed by a sequence of K's (keystrokes). Ignoring these differences and using 1.15 seconds as the time for other number keystrokes, suggests an additional 2.22 seconds (3.37 minus 1.15) was required for the mental operator. As was noted in the literature review, the standard time to mentally prepare is 1.35 seconds (Card, Moran, and Newell, 1983) while Olson and Nilsen (1987 and 1988) estimated a time of 1.62 seconds. The larger value found here reflects situations where the subject was unsure of what to do, violating the routine cognitive task assumption of the model. If a value of 3.25 seconds is used for the initial keystroke time (the approximate mean of digits one through four), the estimate for M is only slightly less, 2.1 seconds, and still larger than values in the literature.

The one estimate not readily obtained from the data set is that for "other" keys because they were only used when there were excess keystrokes. Ignoring that constraint still only leaves 12 data points with a mean time of 3.48 seconds. It is presumed that mean time has an embedded mental operator.

Summary of the analysis process

Based on these data, a five step process emerged for calculating entry times for these data. The estimates are based on 36 drivers performing two imperfectly learned tasks with a specific device, so there are limitations to the data. However, they do provide a reasonable basis for analysis.

Step 1: Select the base keystroke time

In developing a table method for determining keystroke times, both accuracy and ease of use were considered. In examining the various factors, it became apparent that some of the task differences (entry versus retrieval) were due to confounding with the number of repetitions. For that reason and for ease of calculation, task differences were omitted from the model. For shift and space, the mean time for the two tasks was used. Combining Tables 20 and 21, the following values in Table 22 result.

Table 22. Keystroke times for young drivers.

Key Category	Keystroke 1st	2nd	>2nd
Cursor	1.71	0.69	0.47
Enter	1.55		
Letters	1.54	0.99	
Numbers	1.15	0.47	
Shift	1.46		
Space	0.60		

Step 2: Adjust for the lighting conditions

If the key is one that people search for (enter, letter, or number), decrease the time by six percent for dusk conditions. Increase the time by six percent for night conditions. Since this effect is relatively small, this step can be omitted if quick calculations are needed.

Step 3: Adjust for age

Multiply the time by 1.4 for middle-aged drivers, 2.2 for older drivers. For older drivers, the times for shift and enter will be underestimated. However, since there are only a few of those keystrokes in a typical sequence, this is not a major problem.

Step 4: Use 2.22 seconds for mental time.

Step 5: Compute total time

Insert the appropriate values for K and M in the Keystroke-Level Model and compute the total task time.

DEVELOPMENT OF THE KEYSTROKE-LEVEL MODEL

Three steps were involved in developing keystroke-level models to predict destination entry and retrieval times for the Ali-Scout navigation system.

1. Flowcharts were produced indicating all the steps needed to complete each task.
2. Spreadsheets were generated to calculate the times for each task by adding up the times for the steps identified in the flowcharts.
3. Equations summarizing the calculations in the spreadsheet were formulated.

Step 1 - Production of flowcharts

The first step in producing the flowcharts was to analyze the destination retrieval and entry tasks in depth by examining every reasonable method a user could employ to accomplish each part of each task. Methods described in the manual or accounted for by the design of the Ali-Scout were considered reasonable. One method that was not described in the manual but was considered reasonable was that of pressing the shift key prior to each number key during coordinate entry. For alphanumeric fields, the user was required to press the shift key prior to a number key, but for number-only fields (i.e., the coordinate fields), shifting was optional.

Furthermore, an important distinction was made between steps involving pressing buttons (keystroke operators) and those involving thought (mental operators) in accordance with GOMS terminology. Mental operators include verifying the correctness of what was just typed or making a decision about which key to press next. A third category—zero-time step—was added to help indicate branches in the model that were considered not to involve any processing time on the part of the user. For example, a user would probably not take any time to decide between methods for retrieving a destination if he or she had been using the same method for every trial and may not even have been aware of other methods.

As mentioned earlier, three methods were available for performing destination retrieval (see Figure 6)—scroll, type, and hybrid. The scroll method involved using the up and down arrow keys to move through the list of destinations until the desired name appeared on the display. The type method involved typing enough characters of the destination name to uniquely identify it from all the other destinations in the list. The hybrid method involved typing one or more characters of the destination name and then scrolling until the name appeared on the display.

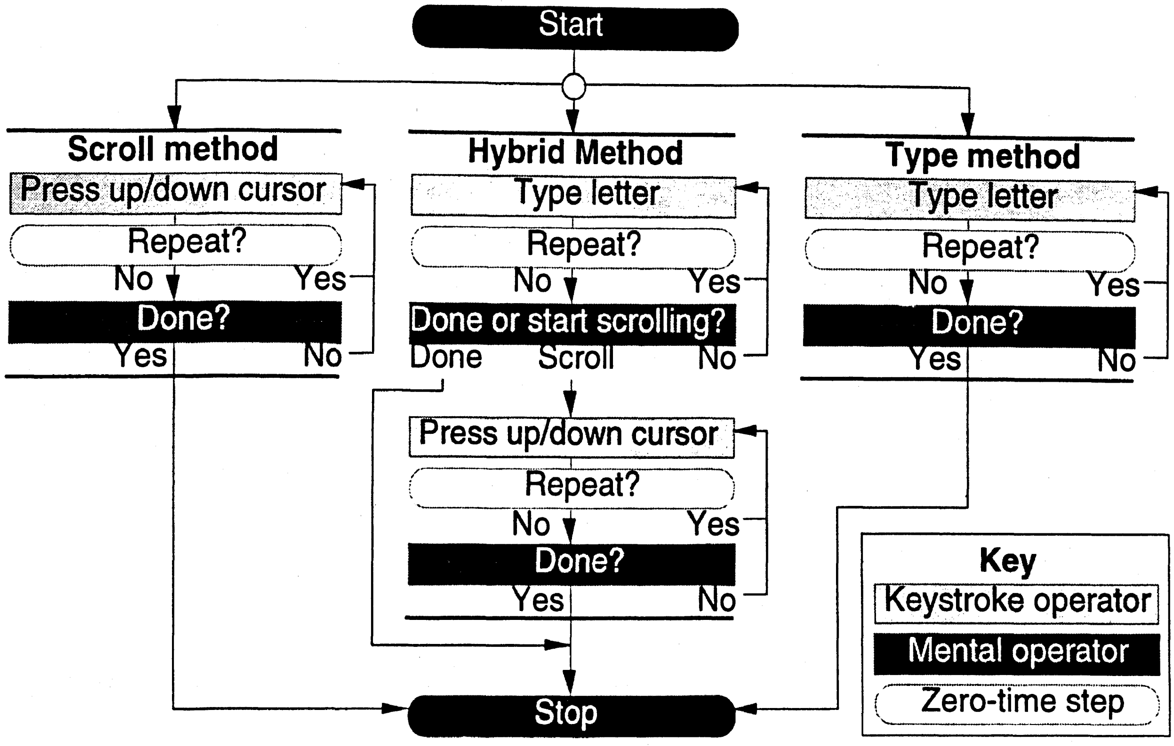


Figure 6. Flowchart for destination retrieval.

In contrast to destination retrieval, there was only one overall method for entering destinations (see Figure 7), although there were several submethods from which the subject could choose as he or she progressed through the task. As the flowchart indicates, the process began with the user typing the entire name of the destination and possibly verifying the spelling once finished. At this point, the user may have advanced to the first coordinate field (longitude) either by pressing enter or the right cursor key. If the cursor key was used, the user must choose a method for moving the cursor from the first to the fourth (last) position of the field since only the last four numbers of each coordinate need to be entered. This could have been accomplished by cursoring over (pressing the right-cursor key until the cursor is in the correct spot) or typing over (typing the first few numbers as they already appeared on the display). Note that the cursor was automatically moved over to the correct spot when the enter key was used.

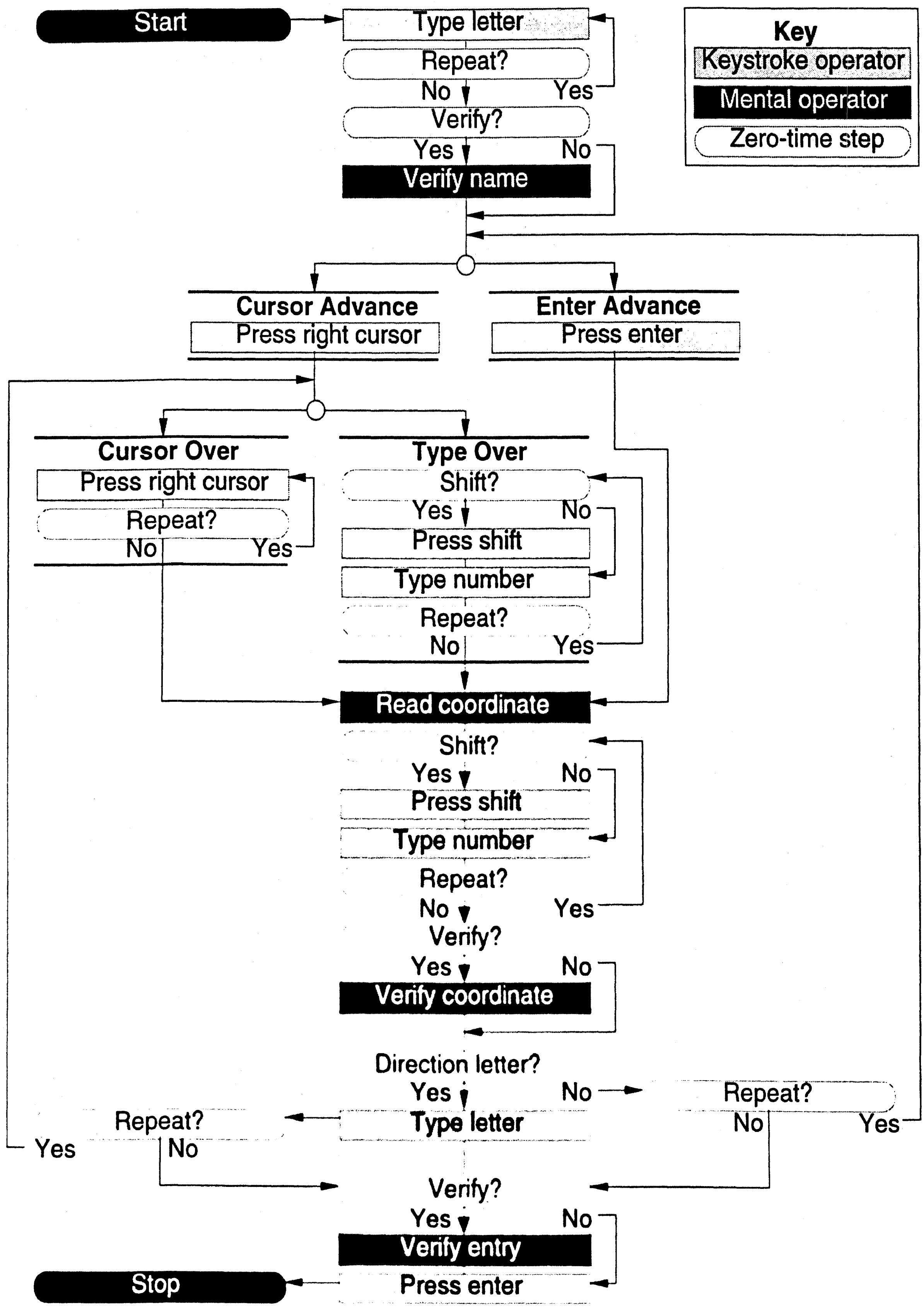


Figure 7. Flow chart for destination entry.

The user must then read the last four digits of the first coordinate from the stimulus card and type them in. The user may choose to press shift before typing each number, and he or she may decide to verify the accuracy of the coordinates once all the digits have been typed. Yet another uncertainty is that the user may decide to type over the direction letter (N) that appears after the last digit of the coordinate. From here, the process is repeated for the second coordinate. Finally, the user must press enter, possibly preceded by a final verification.

The end product of these two flow charts is a series of steps, many of which may or may not occur and many of which are repeated a certain number of times based on the specific destinations to be entered or retrieved. Also, many of the steps may only occur if the user chooses a certain method. Thus, there is a fair amount of branching to the task structure.

Step 2 - Generation of the spreadsheets

A spreadsheet was created for both destination retrieval and entry in which total predicted times were calculated for each age group, lighting condition, and destination card. Each spreadsheet consists of a summary table and several lookup tables. The summary tables contain predicted times for high-level steps and the overall task, while the lookup tables contain adjustment factors for the different age groups and lighting conditions, keystroke (i.e., operator) times, and method probabilities (how often one method was expected to be chosen versus another). The values in the summary table are a function of the information in the lookup tables. Note that some of the values used in the spreadsheets were taken from analyses of the experimental data (described in the results section).

The summary table for the destination retrieval spreadsheet (see Table 23) contains columns for the experimental factors (age, lighting condition, and cards), the three primary methods (scroll, hybrid, and type), and the predicted task time. The task time is based on the time for each method, the probability of each method occurring, and the age adjustment factor. One set of lookup tables (see Tables 24) contains the age and lighting-condition adjustment factors, the keystroke times, and the method probabilities. A second set of lookup tables (see Table 25, 26, and 27) contains the estimated times for each method across all the cards and for both lighting conditions.

Table 23. Portion of the summary table of the retrieval spreadsheet showing the predicted times for the three methods (columns D and E) and the overall task (column G).

	A	B	C	D	E	F	G
1	Age	Condition	Card	Scroll	Hybrid	Type	Predicted Time
2	Young	Dusk	A1	0.773	0.285	0.96507	2.02306667
3	Young	Dusk	A2	0.75733	0.557	4.26853	5.58253333
4	Young	Dusk	A3	0.85917	0.557	3.86027	5.2761
5	...	...	...	...	...	...	...

Table 24. Lookup tables for the retrieval spreadsheet showing adjustment factors, keystroke times, and method probabilities.

	I	J	K	L	M	N	O
1	**Age adjustments**			**Keystroke times**			
2	Age Group	Adjustment*		Operator	1st	2nd	>2nd
3	Young	1		Cursor	1.71	0.69	0.47
4	Middle-aged	1.4		Enter	1.55		
5	Older	2.2		Letter	1.54	0.99	
6	*Applies to overall times			Number	1.15	0.47	
7				Shift	1.46		
8				Space	0.6		
9				Mental	2.22		
10							
11				**Lighting adjustments**			
12				Condition	Adjustment*		
13				Dusk	0.94		
14				Night	1.06		
15				*For enter, letter, or number only			
16							
17				**Probabilities**			
18				Method	Probability		
19				Scroll	0.166666667		
20				Hybrid	0.166666667		
21				Type	0.666666667		

Table 25. Lookup table for the retrieval spreadsheet showing the time for the scroll method by card and lighting condition.

	Q	R	S	T	U	V	W	X
1	**Scroll**							
2	Location Name	Card	Down Cursors	Up Cursors	P(Ideal)*	Cursors	Dusk**	Night**
3	SAKURA BANK	A1	17	6	0.6	10.4	4.638	4.638
4	BIR ICE ARENA	A2	5	18	0.6	10.2	4.544	4.544
5	MONTERREY REST	A3	11	12	0.5	11.5	5.155	5.155
6	MOBIL	A4	9	14	0.5	11.5	5.155	5.155
7	BIG BOY	A5	2	21	0.6	9.6	4.262	4.262
8	SEAFOOD BAY	B1	18	5	0.6	10.2	4.544	4.544
9	PRINT GALLERY	B2	15	8	0.5	11.5	5.155	5.155
10	MAJESTIC CAFE	B3	8	15	0.5	11.5	5.155	5.155
11	VANDENBURG SCH	B4	20	3	0.6	9.8	4.356	4.356
12	BIR LIBRARY	B5	6	17	0.6	10.4	4.638	4.638
13	BILL KNAPPS	C1	3	20	0.6	9.8	4.356	4.356
14	PRIMOS PIZZA	C2	14	9	0.5	11.5	5.155	5.155
15	WOODSIDE HOSP	C3	21	2	0.6	9.6	4.262	4.262
16	BIR THEATER	C4	7	16	0.5	11.5	5.155	5.155
17	MONGOLIAN BBQ	C5	10	13	0.5	11.5	5.155	5.155
18	*Probability of selecting the scroll direction involving the fewest keystrokes							
19	**First cursor not counted because first keystroke of trial was not timed							

Table 26. Lookup table for the retrieval spreadsheet showing the time for the hybrid method by card and lighting condition.

	Q	R	S	T	U	V	W	X	Y
21	**Hybrid**								
22	Location Name	Card	Shifts*	Letters*	1st Curs	2nd Curs	>2nd Curs	Dusk	Night
23	SAKURA BANK	A1	0	0	1	0	0	1.71	1.71
24	BIR ICE ARENA	A2	0	0	1	1	2	3.34	3.34
25	MONTERREY REST	A3	0	0	1	1	2	3.34	3.34
26	MOBIL	A4	0	0	1	1	0	2.4	2.4
27	BIG BOY	A5	0	0	1	0	0	1.71	1.71
28	SEAFOOD BAY	B1	0	0	1	1	0	2.4	2.4
29	PRINT GALLERY	B2	0	0	1	1	1	2.87	2.87
30	MAJESTIC CAFE	B3	0	0	1	0	0	1.71	1.71
31	VANDENBURG SCH	B4	0	1	0	0	0	1.4476	1.6324
32	BIR LIBRARY	B5	0	0	1	1	3	3.81	3.81
33	BILL KNAPPS	C1	0	0	1	1	0	2.4	2.4
34	PRIMOS PIZZA	C2	0	0	1	1	0	2.4	2.4
35	WOODSIDE HOSP	C3	0	0	0	0	0	0	0
36	BIR THEATER	C4	0	0	1	1	4	4.28	4.28
37	MONGOLIAN BBQ	C5	0	0	1	1	1	2.87	2.87
38	*Either letter or shift not counted because first keystroke of trial was not timed								

Table 27. Lookup table for the retrieval spreadsheet showing the time for the type method by card and lighting condition.

	Q	R	S	T	U	V	W
40	**Type**						
41	Location Name	Card	Shifts*	Spaces	Letters*	Dusk	Night
42	SAKURA BANK	A1	0	0	1	1.4476	1.6324
43	BIR ICE ARENA	A2	1	1	3	6.4028	6.9572
44	MONTERREY REST	A3	0	0	4	5.7904	6.5296
45	MOBIL	A4	0	0	2	2.8952	3.2648
46	BIG BOY	A5	0	0	2	2.8952	3.2648
47	SEAFOOD BAY	B1	0	0	3	4.3428	4.8972
48	PRINT GALLERY	B2	0	0	3	4.3428	4.8972
49	MAJESTIC CAFE	B3	0	0	1	1.4476	1.6324
50	VANDENBURG SCH	B4	0	0	1	1.4476	1.6324
51	BIR LIBRARY	B5	1	1	3	6.4028	6.9572
52	BILL KNAPPS	C1	0	0	2	2.8952	3.2648
53	PRIMOS PIZZA	C2	0	0	3	4.3428	4.8972
54	WOODSIDE HOSP	C3	0	0	0	0	0
55	BIR THEATER	C4	1	1	3	6.4028	6.9572
56	MONGOLIAN BBQ	C5	0	0	3	4.3428	4.8972
57	*Either first letter or initial shift not counted because first keystroke						
58	of trial was not timed						

The summary table for the destination entry spreadsheet (see Table 28) contains columns for the experimental factors (age, lighting condition, and cards), the major steps in the process, and the predicted task time. The major steps, which are based on the entry flow chart (see Figure 7 above) consists of (1) typing the name, (2) advancing to the longitude field, (3) moving the cursor over, (4) entering the longitude, (5) advancing to the latitude field, (6) moving the cursor over, (7) entering the latitude, and (8) pressing enter. The task time is based on the time for each step as well as the age adjustment factor. One set of lookup tables (Table 29) contains the age and lighting-condition adjustment factors, the keystroke times, and several groups of method probabilities for both longitude and latitude. A second set of lookup tables (see Tables 30 and 31) contains the estimated times for typing the destination name and entering the longitude and latitude across all the cards and for both lighting conditions.

Table 28. Portion of the summary table of the entry spreadsheet, showing the predicted times for each step (columns D through O) and the overall task (column P).

	A	B	C	D	E	F	G
1	Age	Condition	Card	Type Name	Advance	Move Over	Read Longitude
2	Young	Dusk	A6	15.0884	1.533	0.59155	2.22
3	Young	Dusk	A7	7.8504	1.533	0.59155	2.22
4	Young	Dusk	A8	19.4312	1.533	0.59155	2.22
5	...	...	...	...	...	...	...

	H	I	J	K	L	M
1	Enter Longitude	Direction Letter	Advance	Move Over	Read Latitude	Enter Latitude
2	4.908	0.154	1.438	0.68783	2.22	4.908
3	4.908	0.154	1.438	0.68783	2.22	4.908
4	4.908	0.154	1.438	0.68783	2.22	4.33272
5	...	...	...	...	...	...

	N	O	P
1	Direction Letter	Press Enter	Predicted Time
2	0.154	1.457	35.3594833
3	0.154	1.457	28.1214833
4	0.154	1.457	39.1270033
5	...	...	...

Table 29. Lookup tables for the entry spreadsheet showing adjustment factors, keystroke times, and method probabilities.

Age adjustments

Age Group	Adjustment*
Young	1
Middle-aged	1.4
Older	2.2

*Applies to overall times

Keystroke times

Operator	1st	2nd	>2nd
Cursor	1.71	0.69	0.47
Enter	1.55		
Letter	1.54	0.99	
Number	1.15	0.47	
Shift	1.46		
Space	0.6		
Mental	2.22		

Lighting adjustments

Condition	Adjustment*
Dusk	0.94
Night	1.06

*For enter, letter, or number only

Probabilities

Task	Method	Longitude	Latitude
Type number	Shift, number	0.1	0.1
	Number	0.9	0.9
Advance*	Cursor	0.3	0.5
	Enter	0.7	0.4
Move over**	Cursor	0.25	0.417
	Type	0.05	0.083
Direction Letter	Type	0.1	0.1
	Don't type	0.9	0.9

*Not applicable for lat. if dir. letter is typed for lon.
**Not applicable if enter is used to advance

Table 30. Lookup table for the entry spreadsheet showing the time for typing the name by card and lighting condition.

	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	**Type name**								
2	Location Name	Card	Chars*	1st Letters	2nd Letters	Shifts*	Spaces	Dusk	Night
3	NICKS PLACE	A6	10	9	0	1	1	15.0884	16.7516
4	Q GAS	A7	5	4	0	1	1	7.8504	8.5896
5	HELENS KITCHEN	A8	13	12	0	1	1	19.4312	21.6488
6	YAW GALLERY	A9	10	8	1	1	1	14.5714	16.1686
7	GOODYEAR	A10	7	6	1	0	0	9.6162	10.8438
8	GARMER JACK	B6	10	9	0	1	1	15.0884	16.7516
9	TACO LOCO	B7	8	7	0	1	1	12.1932	13.4868
10	FIRST OF AM	B8	10	8	0	2	2	15.7008	17.1792
11	JACOBSONS	B9	8	8	0	0	0	11.5808	13.0592
12	CHEVRON	B10	6	6	0	1	0	10.1456	11.2544
13	LARK REST	C6	8	7	0	1	1	12.1932	13.4868
14	UNICORN GRILL	C7	12	10	1	1	1	17.4666	19.4334
15	KRGOERS	C8	6	6	0	0	0	8.6856	9.7944
16	QWIK STOP	C9	9	8	0	1	1	13.6408	15.1192
17	TUFFY AUTO	C10	9	7	1	1	1	13.1238	14.5362
18	*Either one character or one shift subtracted because first keystroke of trial was not timed								

Table 31. Lookup tables for the entry spreadsheet showing the time to enter the coordinates, advance the cursor, and move the cursor over.

	Z	AA	AB	AC	AD	AE
20	**Enter longitude**					
21	Coordinate	Card	1st Numbers	2nd Numbers	Dusk	Night
22	1732	A6	4	0	4.908	5.46
23	1654	A7	4	0	4.908	5.46
24	1649	A8	4	0	4.908	5.46
25	1303	A9	4	0	4.908	5.46
26	0508	A10	4	0	4.908	5.46
27	0937	B6	4	0	4.908	5.46
28	1932	B7	4	0	4.908	5.46
29	2459	B8	4	0	4.908	5.46
30	0905	B9	4	0	4.908	5.46
31	1707	B10	4	0	4.908	5.46
32	2250	C6	3	1	4.33272	4.81128
33	0848	C7	4	0	4.908	5.46
34	0506	C8	4	0	4.908	5.46
35	2353	C9	4	0	4.908	5.46
36	1642	C10	4	0	4.908	5.46
37						
38	**Enter latitude**					
39	Coordinate	Card	1st Numbers	2nd Numbers	Dusk	Night
40	3814	A6	4	0	4.908	5.46
41	4038	A7	4	0	4.908	5.46
42	4002	A8	3	1	4.33272	4.81128
43	3302	A9	3	1	4.33272	4.81128
44	3006	A10	3	1	4.33272	4.81128
45	3159	B6	4	0	4.908	5.46
46	3750	B7	4	0	4.908	5.46
47	4307	B8	4	0	4.908	5.46
48	4054	B9	4	0	4.908	5.46
49	3032	B10	4	0	4.908	5.46
50	3236	C6	4	0	4.908	5.46
51	2919	C7	4	0	4.908	5.46
52	2923	C8	4	0	4.908	5.46
53	2936	C9	4	0	4.908	5.46
54	2737	C10	4	0	4.908	5.46
55						
56	**Advance**					
57	Field	Dusk	Night			
58	Longitude	1.5329	1.6631			
59	Latitude	1.4378	1.5122			
60						
61	**Move over**					
62	Field	Cursor	Type (Dusk)	Type (Night)	Dusk	Night
63	Longitude	1.63	3.681	4.095	0.59155	0.61225
64	Latitude	1.16	2.454	2.73	0.687833333	0.710833333

Step 3 - Formulation of equations

Equations for calculating the total predicted time for destination retrieval and entry were generated by extracting the appropriate spreadsheet formulas and replacing the cell references (e.g., A5) with more meaningful symbols. The symbols (see Table 32) are in the form $Element_{object}$, where the element can be a time, number, probability, or adjustment factor and the object can be a task, method, operator, or variable. For instance, T_{entry} refers to the time to enter a single destination while A_{light} is a multiplier whose value varies according to the lighting condition.

Table 32. The coding scheme used for the equations.

Element	Symbol	Description
Time	T_x	Time for task, method, or operator *x*
Number	N_x	Number of times method or operator *x* occurs
Probability	P_x	Probability of method or operator *x* occurring
Adjustment factor	A_x	Adjustment factor for variable *x*

The total retrieval time (see Table 33) is calculated by multiplying the time for each method (scroll, hybrid, and type) by the probability of that method occurring, then taking the sum of these products, and, finally, multiplying this sum by the age adjustment factor. The time for the scroll method is a function of the number of scrolls, which, in turn, is a function of the probability that the user chooses the optimal scroll direction.

Table 33. Equations for calculating the total retrieval time.

Eq #	Equation		
1	$T_{retrieval}$	=	$A_{age} * (P_{scroll}T_{scroll} + P_{hybrid}T_{hybrid} + P_{type}T_{type})$
1.1	T_{scroll}	=	$T_{2nd\ cursor} + (N_{scrolls} - 2) * T_{>2nd\ cursor}$
1.2	T_{hybrid}	=	$N_{shifts\ (hybrid)}T_{shift} + A_{light}N_{letters\ (hybrid)}T_{letter} + N_{1st\ cursors}T_{1st\ cursor} + N_{2nd\ cursors}T_{2nd\ cursor} + N_{>2nd\ cursors}T_{>2nd\ cursor}$
1.3	T_{type}	=	$N_{shifts\ (type)}T_{shift} + N_{spaces}T_{space} + A_{light}N_{letters}T_{letter}$
1.1.1	$N_{scrolls}$	=	$P_{ideal} * \min(N_{up\ cursors}, N_{down\ cursors}) + (1 - P_{ideal}) * \max(N_{up\ cursors}, N_{down\ cursors})$
1.1.1.1	P_{ideal}	=	the probability that the user chooses the fastest direction to scroll

Note: These equations do not take into account the first keystroke of the task since this was not timed during the experiment.

The equations for the total entry time appear in Table 34. This time is calculated by adding up the times for each of 12 steps (see equations 1.1 through 1.12) and multiplying this sum by the age adjustment factor. Moving the cursor over within the longitude and latitude fields is accomplished either by cursoring over or typing over.

Table 34. Equations for calculating the total entry time.

Eq #	Equation
1	$T_{entry} = A_{age} * (1.1 + 1.2 + 1.3 + ... + 1.10 + 1.11 + 1.12)$
1.1	$T_{type\ name} = A_{light}N_{1st\ letters}T_{1st\ letters} + A_{light}N_{2nd\ letters}T_{2nd\ letter} + N_{shifts}T_{shift} + N_{spaces}T_{space}$
1.2	$T_{advance\ (lon)} = P_{curs\ advance\ (lon)}T_{1st\ curs} + A_{light}P_{enter\ advance\ (lon)}T_{enter}$
1.3	$T_{move\ over\ (lon)} = P_{curs\ over\ (lon)}T_{curs\ over\ (lon)} + P_{type\ over\ (lon)}T_{type\ over\ (lon)}$
1.4	$T_{read\ (lon)} = T_{mental}$
1.5	$T_{enter\ (lon)} = P_{shift,\ num} * (4T_{shift} + A_{light}N_{1st\ nums\ (lon)}T_{1st\ num} + A_{light}N_{2nd\ nums\ (lon)}T_{2nd\ num}) + P_{num} * (A_{light}N_{1st\ nums\ (lon)}T_{1st\ num} + A_{light}N_{2nd\ nums\ (lon)}T_{2nd\ num})$
1.6	$T_{dir\ letter\ (lon)} = P_{dir\ letter\ (lon)}T_{1st\ letter}$
1.7	$T_{advance\ (lat)} = P_{curs\ advance\ (lat)}T_{1st\ curs} + A_{light}P_{enter\ advance}T_{enter\ advance\ (lat)}T_{enter}$
1.8	$T_{move\ over\ (lat)} = P_{curs\ over\ (lat)}T_{curs\ over\ (lat)} + P_{type\ over\ (lat)}T_{type\ over\ (lat)}$
1.9	$T_{read\ (lat)} = T_{mental}$
1.10	$T_{enter\ (lat)} = P_{shift,\ num} * (4T_{shift} + A_{light}N_{1st\ nums\ (lat)}T_{1st\ num} + A_{light}N_{2nd\ nums\ (lat)}T_{2nd\ num}) + P_{num} * (A_{light}N_{1st\ nums\ (lat)}T_{1st\ num} + A_{light}N_{2nd\ nums\ (lat)}T_{2nd\ num})$
1.11	$T_{dir\ letter} = P_{dir\ letter\ (lat)}T_{1st\ letter}$
1.12	$T_{press\ enter} = T_{enter}$
1.3.1	$T_{curs\ over\ (lon)} = T_{2nd\ curs} + 2T_{>2nd\ curs}$
1.3.2	$T_{type\ over\ (lon)} = P_{shift,\ num} * 3 * (T_{shift} + A_{light}T_{1st\ num}) + P_{num}A_{light}T_{1st\ num}$
1.8.1	$T_{curs\ over\ (lat)} = T_{2nd\ curs} + T_{>2nd\ curs}$
1.8.2	$T_{type\ over\ (lat)} = P_{shift,\ num} * 2 * (T_{shift} + A_{light}T_{1st\ num}) + P_{num}A_{light}T_{1st\ num}$

Note: These equations do not take into account the first keystroke of the task since this was not timed during the experiment.

lon = longitude, lat = latitude

EVALUATION OF THE KEYSTROKE-LEVEL PREDICTIONS

Five sets of predictions based on the keystroke-level analyses were compared with the trial times from the experiment. These included:

1. A standard keystroke-level analysis (as implemented in Kieras, 1988) where K was assumed to be 1.2 seconds (worst typist) and M was 1.35 seconds.
2. A modified analyses using values of K and M more typical of this experiment (1.5 and 2.2 seconds respectively).
3. A standard keystroke-level analysis (K=1.2, M=1.35) with adjustments for age (multiple by 1.4 for middle-aged subjects, 2.2 for older subjects).
4. A tailored keystroke-level analysis that had unique K values for each category, adjustments for repetitions, adjustments for lighting, but no adjustments for age.
5. A tailored keystroke-level analysis that had unique K values for each category, adjustments for repetitions, adjustments for lighting, and adjustments for age.

For each set of predictions, regression plots were generated for both entry and retrieval, for both young subjects (the best case) and all subjects. Figures 21 and 22 show the results for the best of the five sets, the tailored keystroke-level model. (See Appendix A for the full set of plots.) Correlations ranged from 0.70 to 0.91, with those for retrieval being higher than those for entry, and those for young subjects being higher than those for all subjects. Readers should keep in mind that the actual times are not the true times, but only experimental estimates. From that viewpoint, the correlations reported are quite good. The higher correlations for the retrieval task may suggest that it had become more routine than the entry task, especially among young subjects. Since one of the assumptions of the GOMS models is that the task be routine, it makes sense that the model for the task with fewer steps to learn and master would be more accurate. Furthermore, the ability to build successful keystroke-level models may also be a factor of the total time for the task, with shorter tasks being easier to predict than longer ones. Finally, it is no surprise that predictions for young subjects were more reliable than those for all subjects pooled together since younger people tend to learn new tasks more quickly and perform them more fluidly.

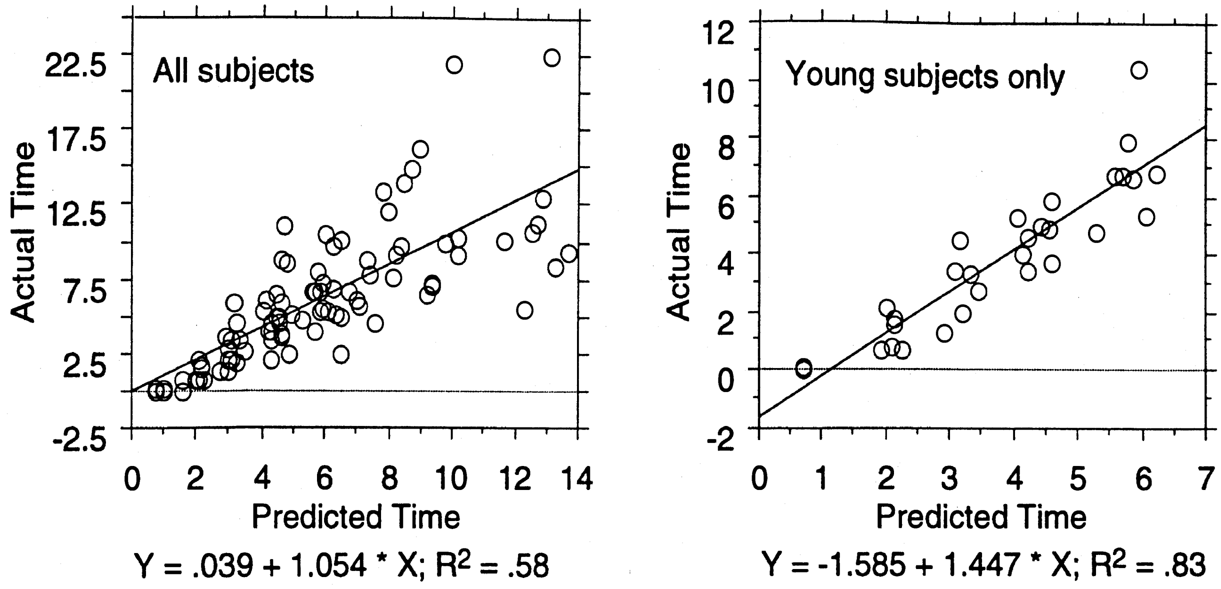


Figure 21. Actual versus predicted times for destination retrieval using all the data (left) and only the data from young subjects (right).

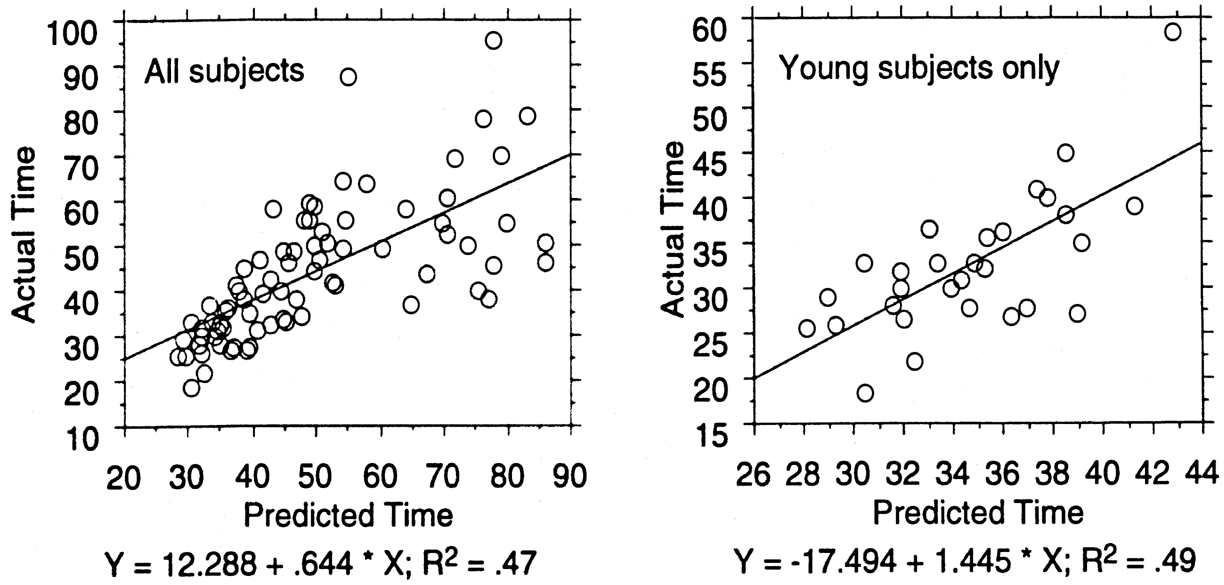


Figure 22. Actual versus predicted times for destination entry using all the data (left) and only the data from young subject (right).

Table 35 shows the regression equations and predictive equations for all combinations examined. The extent to which particular factors predicted performance varied quite widely from situation to situation. For example, the standard keystroke-level predictions accounted for 78 percent of the variance of actual retrieval times for young subjects, a fairly high value. However, when all subjects were included, the value dropped to 41 percent. For entry by young subjects, 38 percent of the variance was accounted for, while the value for all subjects was quite low, 12 percent.

Table 35. Regression equations and correlations for various models.

Retrieval	Young		All Subjects	
	Equation	R2	Equation	R2
standard Keystroke-Level Model (K=1.2, M=1.35)	Y=-2.381 + 1.380 X	0.784	Y=-1.839 + 1.759 X	0.414
modified keystroke model (K=1.5, M=2.2)	Y=-2.381 + 1.104 X	0.784	Y=-1.839 + 1.407 X	0.414
standard model (K=1.2, M=1.35) with age adjustment (young=1, middle=1.4, old=2.2)			Y=-0.352 + 0.935 X	0.540
tailored model (various K's, M=2.2, lighting values, no age adjustments)			Y=-0.853 + 1.851 X	0.443
tailored model (various K's, M=2.2, lighting values, age adjustments)	Y=-1.585 + 1.447 X	0.832	Y=0.039 + 1.054 X	0.583
Entry				
standard Keystroke-Level Model (K=1.2, M=1.35)	Y=-17.057 + 1.613 X	0.385	Y=-12.962 + 1.851 X	0.123
modified keystroke model (K=1.5, M=2.2)	Y=-18.379 + 1.290 X	0.385	Y=-14.479 + 1.480 X	0.123
standard model (K=1.2, M=1.35) with age adjustment (young=1, middle=1.4, old=2.2)			Y=13.57 + 0.692 X	0.454
tailored model (various K's, M=2.2, lighting values, no age adjustments)			Y= -7.689 + 1.498 X	0.119
tailored model (various K's, M=2.2, lighting values, age adjustments)	Y=-17.494 + 1.445 X	0.486	Y=12.288 + 0.644 X	0.472

Y is the actual time, X is the prediction.

Notice that using keystroke-level values based on the experiment (K=1.5, M=2.2) does not improve the correlation, but it does reduce the size of the adjustment necessary to fit the actual data. Depending on the condition, the multiplier drops from about 1.4 to 1.9 to 1.1 to 1.5. In all cases, some adjustment to the intercept is also necessary, complicating the adjustment. These adjustments should be made to keystroke-level estimates where differences in the relative or absolute size of alternative interfaces are of interest.

Another important insight from the data is that the primary adjustment needed is for subject age. For example, for all subjects, adding adjustments for age to the basic keystroke-level model raised the R2 from 0.41 to 0.54 whereas adding adjustments for everything but age (lighting, Ks tailored for each character group, repetitions, a revised M) only raised the correlation to 0.44. To put these values in perspective, the best model (all factors included) had an R2 of 0.58 and needed minimal additional adjustments to predict actual times. Thus, these data highlight the importance of adjustments for age. For quick hand calculations, there may not be much value in using tailored values to adjust for the specific key type or repetition effects, if rank orders of alternative interfaces are desired. For exact estimates of alternatives, the tailored keystroke-level models gave the best predictions and required the fewest post hoc adjustments to predict actual times, especially for retrieval tasks.

CONCLUSIONS

Overall, these data suggest that a keystroke-based model at the keystroke level can be used to provide reasonable estimates of driver performance in destination entry and retrieval when these tasks are performed alone. However, estimates based on textbook data are likely to grossly underestimate real world performance. New values are required for model parameters and additional adjustments are needed to provide the desired accuracy as described below.

What are typical keystroke and mental operator times?

The time per keystroke depends on the particular character typed. While this is quite evident in the typing literature (Evey, 1980), these differences are ignored in traditional keystroke-level analyses. For young subjects, mean interkeystroke intervals were about 1.7 seconds for initial cursor keystrokes, 1.5 seconds for initial enter, letters, and shift keystrokes, 1.15 seconds for numbers, and 0.6 seconds for space. These times are consistently larger than those in the GOMS literature (0.5 s/keystroke for typing random letters, 0.7 s/keystroke for complex codes, or 1.2 s/keystroke for a worst typist). Larger values were obtained here due to poor keyboard tactile feedback, small key size, and delays due to mental activity (as the keying task was not fully learned).

Repetition effects are well known, but previously they have been ignored in keystroke-level analyses. Depending on the key type, a second repeated keystroke requires about half (plus or minus 10 percent) of the time for the initial keystroke. Subsequent keystrokes are about 30 percent of the time for the initial keystroke. Specifically, second keystrokes were about 1 second for letters, 0.7 seconds for cursor actions, and 0.5 seconds for numbers, times comparable to those reported in the literature for typing complex codes (.75 s/keystroke). For more than 2 cursor keystrokes, times were about 0.5 seconds.

Keystroke times were also affected by lighting, but the impact of lighting was less pronounced than for other factors (though basically only a limited range of dark conditions were examined). Baseline times should be increased by six percent for night conditions, decreased by six percent for dusk.

Age differences were large and depended on the key typed. On average, middle-aged drivers took 1.45 times longer than younger drivers, while for older drivers, the difference was a factor of 2.19. However, there were some differences between key types. For example, shift key operations took much longer for older subjects, probably because the keypress included an embedded mental operator. This suggests that for older drivers for some tasks, additional mental operators need to be added to predictions. For simplicity, it is recommended that times should also be adjusted for age, multiplying times for young subjects by 1.4 to estimate performance for middle-aged subjects and 2.2 for older subjects.

Not only do textbook values underestimate keying times, but mental times are underestimated as well. Mental times were about 2.2 seconds, a value 62 percent greater than those reported in the literature (1.35 s).

How well did GOMS models predict overall times?

R2 from equations linking predicted and actual times ranged from approximately 0.1 to 0.8 for the unmodified keystroke-level model and 0.5 to 0.8 for the tailored model. Readers are again reminded that the actual values are really estimates of the true times, and from that perspective, the models were accounting for a significant fraction of the total variance. Further, the entry and retrieval tasks were not fully learned in all cases, that is, executing them did not only involve routine cognitive skill. However, this is true of many of the tasks that drivers will perform with in-vehicle information systems. Procedures are needed to determine where mental processes may sometimes occur and the time required for them.

Correlations were greater for younger drivers than for older drivers and greater for retrieval than for destination entry. Of the various improvements to the basic model explored (tailored Ks and Ms, adjustments for lighting, etc.), the single most important adjustment was for age, with the age multipliers alone (1.4 for middle-aged drivers, 2.2 for older drivers as was noted earlier) leading to basic keystroke-level predictions that were almost as good as the tailored model. Further adjustments required use of two parameter corrections (slope and intercept) to the predictions.

Thus, overall these data suggest that reasonable predictions of task completion time can be obtained from keystroke-level estimates of driver performance. However, these data suggest, that using age adjustments (that may be task specific) are desired. For exact predictions, tailored estimates of M and K may also be desired. These data suggest using a larger value of M than is reported in the literature, and that key type differences, driver age, repetition effects, and lighting conditions may warrant consideration.

As was noted at the outset of this report, manufacturers of equipment often do not have the time or resources to experimentally evaluate every user interface or interface modification they produce. Accordingly, there are many potential opportunities for keystroke-level models. This report provides data for improving the accuracy of keystroke-level estimates, especially for applications where the task has yet to become ingrained and the interface is far from ideal (the real world). While the estimates are highly parameterized to the specific tasks investigated, the authors believe the tasks examined are similar to those in a wide variety of practical contexts. Further, the rules of thumb developed for correcting for age and key repetitions should also be generally useful. Readers who apply these estimates to other interfaces are encouraged to contact the authors with their results.

REFERENCES

Barbaresso, J.C. (1994). Preliminary Findings and Lessons Learned from the FAST-TRAC IVHS Program, Proceedings of the IVHS America 1994 Annual Meeting, volume 1, Washington, D.C.: IVHS America, 489-497.

Beevis, D. and St. Denis, G. (1992). Rapid Prototyping and the Human Factors Engineering Process, Applied Ergonomics, 23(3), 155-160.

Berger, C., Walton, C., and Wurman, P. (1993). The Event Recorder (version 5.2), Ann Arbor, MI: The University of Michigan, Office of Instructional Technology.

Card, S.K., Moran, T.P., and Newell, A. (1980). The Keystroke-Level Model for User Performance Time with Interactive Systems, Communications of the ACM, July, 23(7), 396-410.

Card, S.K., Moran, T.P., and Newell, A. (1983). The Psychology of Human-Computer Interaction, Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, W.E. (ed). (1983). Cognitive Aspects of Skilled Typewriting, Berlin, Germany: Springer-Verlag.

Detweiler, M.C. (1990). Alphabetic Input on a Telephone Keypad, Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA: Human Factors Society, 212-216.

Eby, D.W., Streff, F.M., Wallace, R.R., Kostyniuk, L.P., Hopp, M.L., and Underwood, S. (1996). An Evaluation of User Perceptions and Behaviors of FAST-TRAC: Pilot Study Results (Technical Report UMTRI-96-14), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Evey, R.J. (1980). How Typists Type (IBM Technical Report HFC-35), San Jose, CA: IBM General Products Division, Human Factors Center.

Green, P., Boreczky, J., and Kim, S. (1990). Applications of Rapid Prototyping to Control and Display Design. (SAE paper #900470, Special Publication SP-809), Warrendale, PA: Society of Automotive Engineers.

Green, P., Levison, W., Paelke, G., and Serafin, C. (1995). Preliminary Human Factors Guidelines for Driver Information Systems (Technical Report FHWA-RD-94-087), McLean, VA: U.S. Department of Transportation, Federal Highway Administration.

Green, P. and Olson, A. (1997). A Technical Description of the UMTRI Driving Simulator Family-1996 Implementation (Technical Report UMTRI 97-12), Ann Arbor, MI: University of Michigan Transportation Research Institute (in preparation).

Haunold, P. and Kuhn, W. (1994). A Keystroke Level Analysis of a Graphics Application: Manual Map Digitizing, CHI'94 Proceedings, New York: Association for Computing Machinery, 337-343.

Katz, S., Fleming, J., Hunter, D.R., Green, P., and Damouth, D. (1996). On-the-Road Human Factors Evaluation of the Ali-Scout Navigation System (Technical Report UMTRI-96-32), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Katz, S., Green, P. , and Fleming, J. (1995). Calibration and Baseline Driving Data for the UMTRI Driver Interface Research Vehicle, (Technical Report UMTRI-95-2), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Kieras, D. (1988). Towards a Practical GOMS Model Methodology for User Interface Design, Chapter 7 in M. Helander (ed.) Handbook of Human-Computer Interaction, New York: Elsevier Science, 135-157.

Kostyniuk, L. and Eby, D. (1996). Natural Use/Yoke Study Survey (Technical Report EECS-ITS LAB-FT95-018), Ann Arbor, MI: The University of Michigan Intelligent Transportation Systems Center.

MacAdam, C.C., Green, P.A., and Reed, M.P. (1993). An Overview of Current UMTRI Driving Simulators, UMTRI Research Review, July-August, 24(1), 1-8.

Manes, D., Green, P., and Hunter, D. (1996). Glance Frequencies to the Ali-Scout Navigation System (Technical Report UMTRI-96-42), Ann Arbor, MI: The University of Michigan Transportation Research Institute (in preparation).

Manes, D., Green, P., and Hunter, D. (1996). Prediction of Destination Entry and Retrieval Times Using GOMS (Technical Report UMTRI-96-37), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Olson, A. and Green, P. (1997). A Description of the UMTRI Driving Simulator Architecture and Alternatives (Technical Report UMTRI-97-15), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Paelke, G.M. (1993). A Comparison of Route Guidance Destination Entry Methods, Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting-1993, Santa Monica, CA: The Human Factors and Ergonomics Society, 569-573.

Paelke, G. and Green, P. (1993). Entry of Destinations into Route Guidance Systems: A Human Factors Evaluation (Technical Report UMTRI-93-45), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Siemens Automotive (undated). Ali-Scout Navigation System User's Guide, Auburn Hills, MI: Siemens Automotive.

Segal, L.D. and Andre, A.D. (1993). Activity Catalog Tool (A.C.T.) v 2.0 User Manual (NASA Contractor Report CR 177634), Moffett Field, CA: NASA Ames Research Center.

Steinfeld, A., Manes, D., Green, P., and Hunter, D. (1996). Destination Entry and Retrieval with the Ali-Scout Navigation System (Technical Report UMTRI-96-30), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Taylor, W. and Wu, J. (1995). A Database System Containing MOE's of Interest to the Evaluation (Technical Report EECS-ITS LAB FT95-028), Ann Arbor, MI: The University of Michigan Intelligent Transportation Center.

Treece, J.B. (1996). In Japan, Car Buyers Put Navigation Devices at the Top of Their List, Automotive News, September 16, 1996, 3 & 24.

Underwood, S.E. (1994). FAST-TRAC: Evaluating an Integrated Intelligent Vehicle-Highway System, Proceedings of the IVHS America 1994 Annual Meeting, 1, Washington, D.C.: IVHS America, 300-311.

Verwey, W.B. (1996). Buffer Loading and Chunking in Sequential Keypressing, Journal of Experimental Psychology: Human Perception and Performance, 22(3), 544-562.

Zwahlen, H.T., Adams, C.C. Jr., and DeBald, D.P. (1988). Safety Aspects of CRT Touch Panel Controls in Automobiles, in Gale, A.G., Freeman, M.H., Haslegrave, C.M., Smith, P., and Taylor, S.P., Vision in Vehicles II, Amsterdam, Netherlands: Elsevier Science, 335-344.

Zwahlen, H.T. and DeBald, D.P. (1986). Safety Aspects of Sophisticated In-Vehicle Information Displays and Controls. Proceedings of the Human Factors Society-30th Annual Meeting, Santa Monica, CA: Human Factors Society, 256-260.

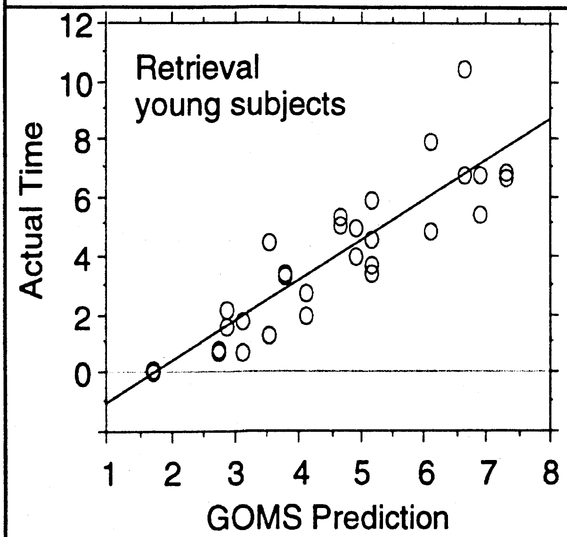
APPENDIX A - REGRESSION PLOTS OF VARIOUS KEYSTROKE-LEVEL MODELS

Note: In the figures and tables that follow, GOMS is used as a shorthand to indicate variations of the keystroke-level model.

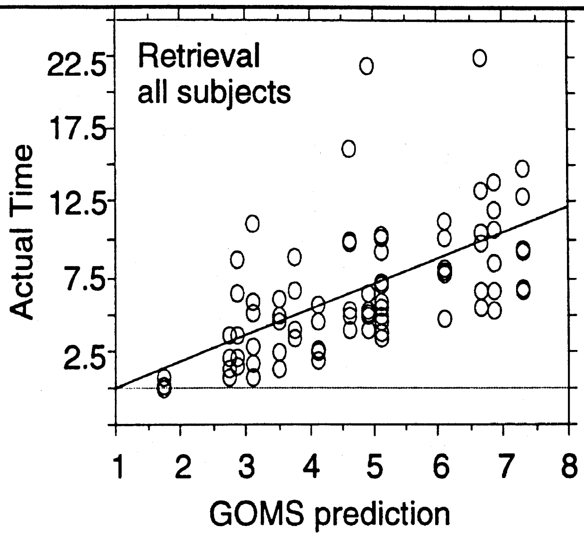
RETRIEVAL

Young	All Subjects

Basic keystroke-level model (K=1.2, M=1.35)

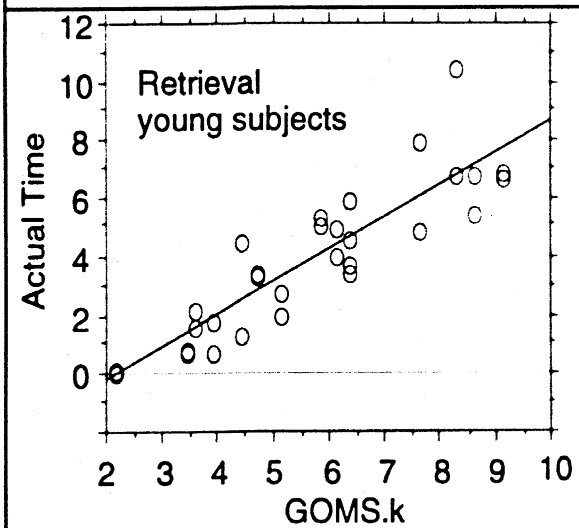


Actual Time = -2.381 + 1.38 * GOMS Prediction; R^2 = .784

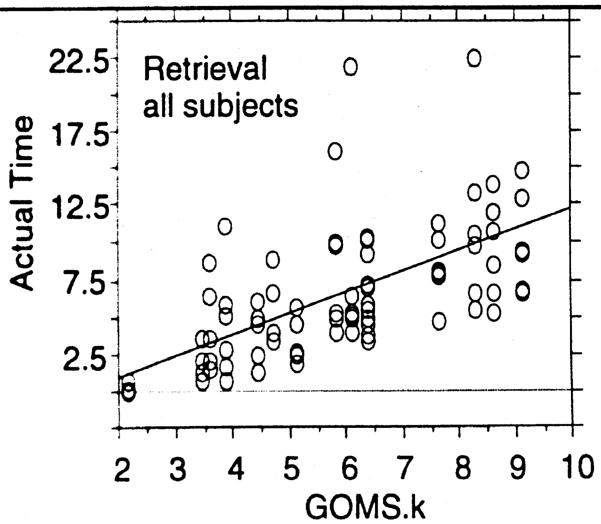


Actual Time = -1.839 + 1.759 * GOMS
R^2 = .414

Improved K & M (K=1.5, M=2.2)



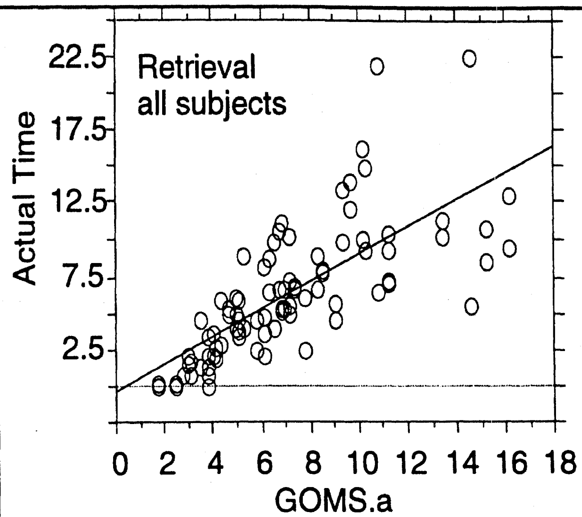
Actual Time = -2.381 + 1.104 * GOMS.k
R^2 = .784



Actual Time = -1.839 + 1.407 * GOMS.k
R^2 = .414

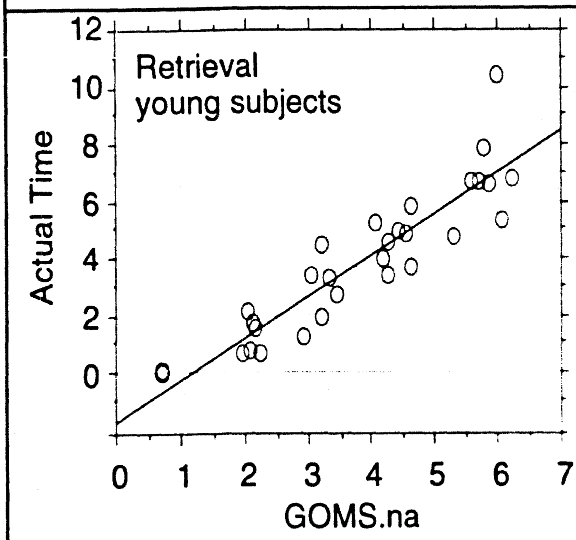
Keystroke model with age adjustment (y,mid,old=1.0, 1.4, 2.2)

same as basic GOMS for retrieval without age adjustment

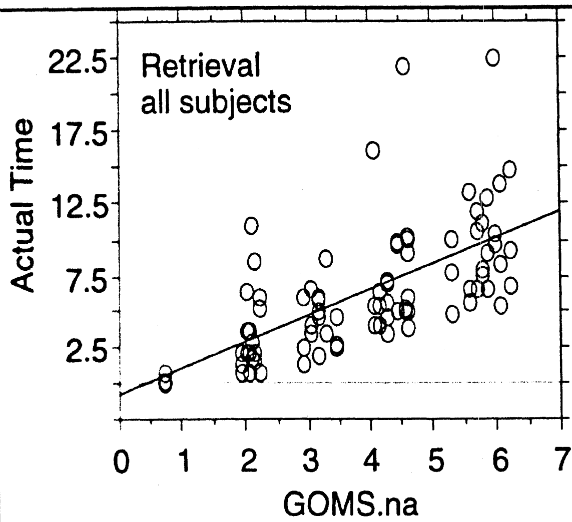


Actual Time = -.352 + .935 * GOMS.a
R^2 = .54

Tailored model no age adjustment (various K's, M=2.2, lighting adjust.)

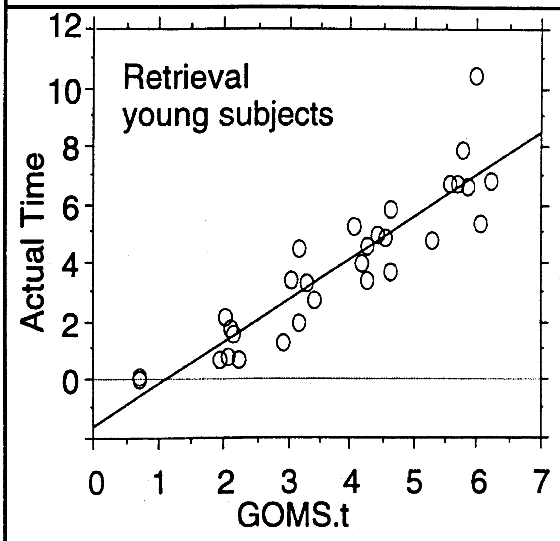


Actual Time = -1.585 + 1.447 * GOMS.na
R^2 = .832

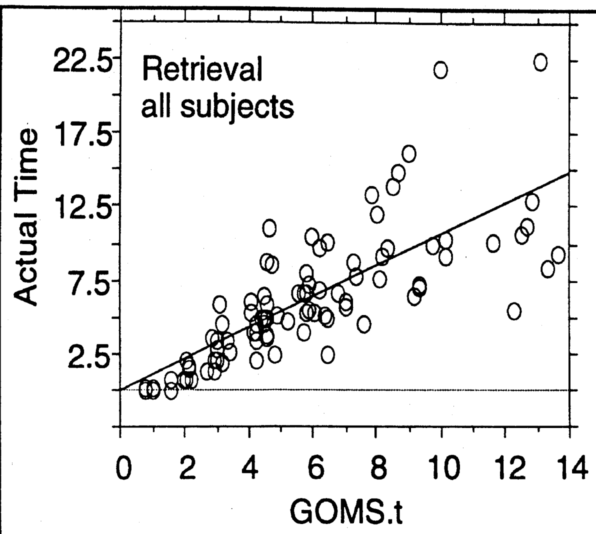


Actual Time = -.853 + 1.851 *GOMS.na
R^2 = .443

Tailored model with age adjustment (various K's, M=2.2, lighting adjust.)



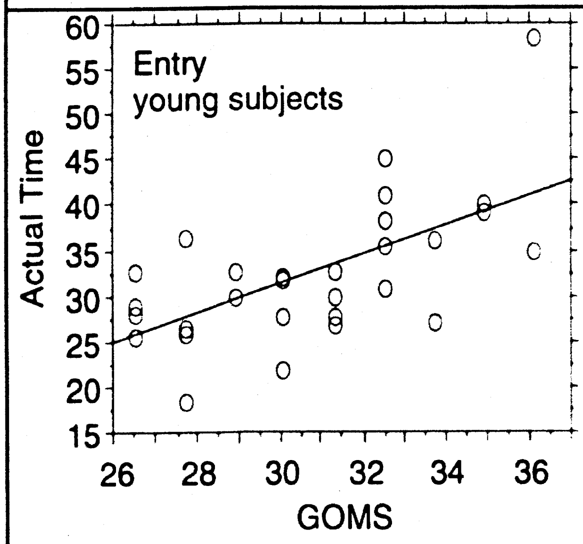
Actual Time = -1.585 + 1.447 * GOMS.t
R^2 = .832



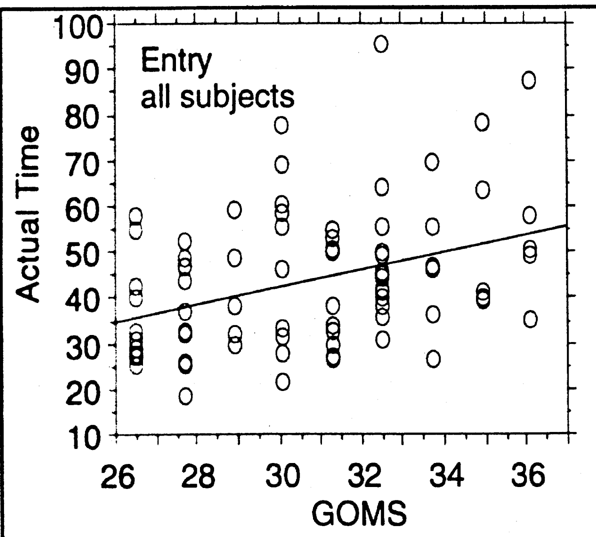
Actual Time = .039 + 1.054 *GOMS.t
R^2 = .583

ENTRY

Basic keystroke-level model (K=1.2, M=1.35)

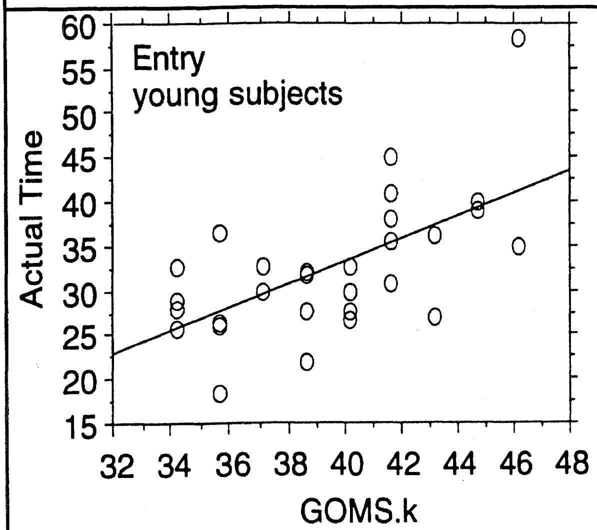


Actual Time = -17.057 + 1.613 * GOMS
R^2 = .385

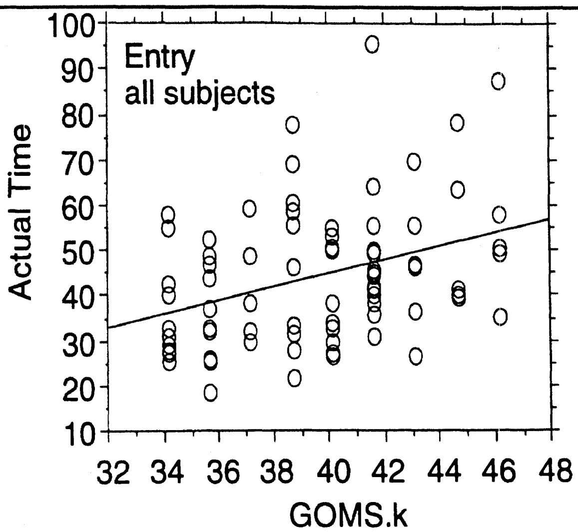


Actual Time = -12.962 + 1.851 * GOMS
R^2 = .123

Improved K & M (K=1.5, M=2.2)



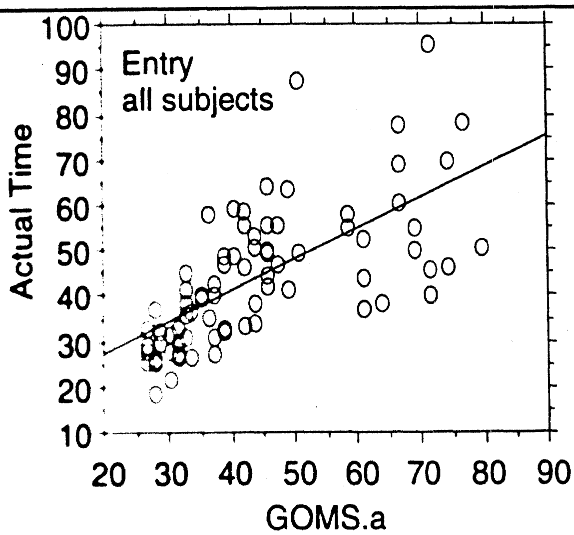
Actual Time = -18.379 + 1.29 *GOMS.k
R^2 = .385



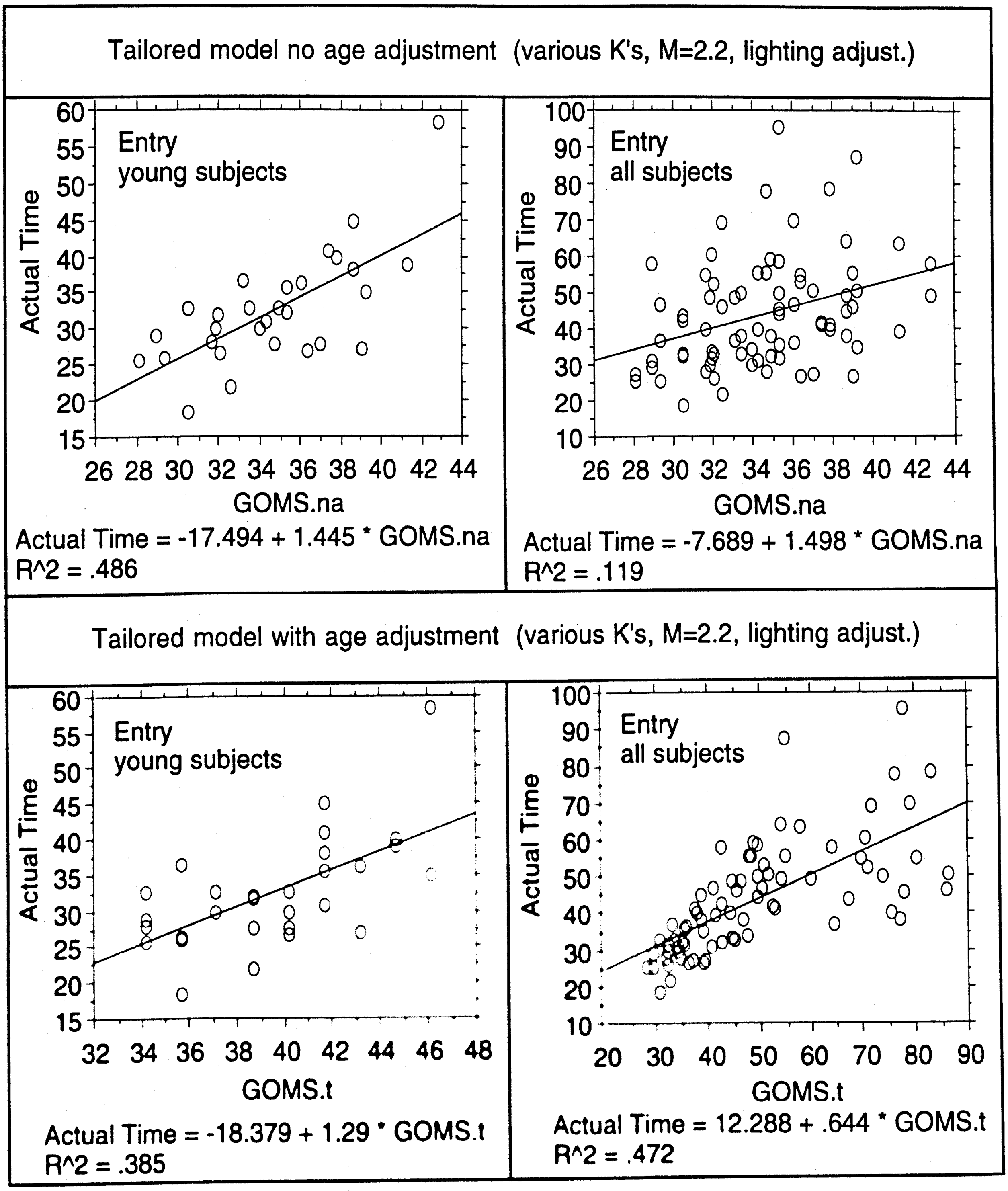
Actual Time = -14.479 + 1.48 * GOMS.k
R^2 = .123

Keystroke model with age adjustment (y,mid,old=1.0, 1.4, 2.2)

same as basic GOMS for entry without age adjustment



Actual Time = 13.57 + .692 * GOMS.a
R^2 = .454


Tailored model no age adjustment (various K's, M=2.2, lighting adjust.)
Entry
young subjects
Actual Time
GOMS.na
Actual Time = -17.494 + 1.445 * GOMS.na
R^2 = .486
Entry
all subjects
Actual Time
GOMS.na
Actual Time = -7.689 + 1.498 * GOMS.na
R^2 = .119
Tailored model with age adjustment (various K's, M=2.2, lighting adjust.)
Entry
young subjects
Actual Time
GOMS.t
Actual Time = -18.379 + 1.29 * GOMS.t
R^2 = .385
Entry
all subjects
Actual Time
GOMS.t
Actual Time = 12.288 + .644 * GOMS.t
R^2 = .472