90529

Technical Report UMTRI-97-42 November, 1997

Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth)

Daniel Manes and Paul Green

HUMAN FACTORS

UMTRI The University of Michigan Transportation Research Institute

Technical Report Documentation Page

1. Report No.
UMTRI-97-42

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth)

5. Report Date
November, 1997

6. Performing Organization Code
account 374912

7. Author(s)
Daniel Manes and Paul Green

8. Performing Organization Report No.
UMTRI-97-42

9. Performing Organization Name and Address
The University of Michigan
Transportation Research Institute (UMTRI)
2901 Baxter Rd, Ann Arbor, Michigan
48109-2150 USA

10. Work Unit no. (TRAIS)

11. Contract or Grant No.
DRDA 97-1142

12. Sponsoring Agency Name and Address
United Technologies Automotive
Advanced Technologies Development
5200 Auto Club Drive
Dearborn, MI 48126-2659 USA

13. Type of Report and Period Covered
1/97 - 11/97

14. Sponsoring Agency Code

15. Supplementary Notes
Project Liaison: Mr. Tim Smith

In an initial experiment, four drivers were verbally cued to find items in a menu system while driving a simulator. Three factors were examined: (1) menu structures (deep with three levels of four items each and, broad with two levels of eight items each), (2) controls (a cursor and a number pad), and (3) control/display locations (both high, both low, and display high/control low). The cursor control led to faster task completion times than the number pad, and a full 32 trials per block offered no benefit over just 16, simplifying the design of the main experiment.

The main experiment was similar to the initial study, except 24 drivers participated, a knob control was used instead of the number pad, and 16 rather than 32 trials per block were used. Menu selection times (approximately nine seconds) were not significantly different for the two menu structures or the three control/display locations. However, the knob was significantly faster than the cursor (six percent), and there were 13 percent more lane excursions for the broad menus than the deep.

A GOMS analysis showed that downward scrolls required 0.4 seconds each, except for initial selections where thinking and reading processes lengthened times. Correlations between GOMS and actual selection times ranged from 0.5 to 0.8.

These data, suggest deep menus should be used to minimize lane excursions (a measure of driver distraction) and knobs should be utilized to minimize selection times. Mounting the display high and the control low is also advised.

17. Key Words
ITS, human factors, ergonomics, driving, usability, input devices, controls, keyboards, menu, GOMS

18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classify. (of this report)
(None)

20. Security Classify. (of this page)
(None)

21. No. of pages
92

22. Price

Form DOT F 1700 7 (8-72)

Reproduction of completed page authorized

Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth)

UMTRI Technical Report 97-42
Daniel Manes and Paul Green

University of Michigan
Ann Arbor, Michigan, USA

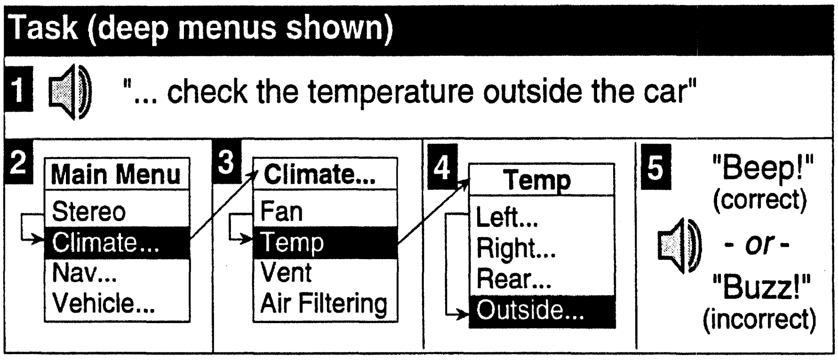
1 ISSUES

1. What are typical selection times & error rates for novice users & can times be predicted using a GOMS model?
2. How do selection times & error rates vary as a function of (a) menu structure, (b) control type, & (c) the location of the control and display?
3. What is the effect of driver age, sex, and practice on selection times & error rates?
4. How acceptable is the idea of an in-vehicle menu interface to drivers?

2 METHOD

 # subjects (n = 24)

Age	M	F
Young (21-27)	6	6
Older (65-70)	6	6



Task (deep menus shown)

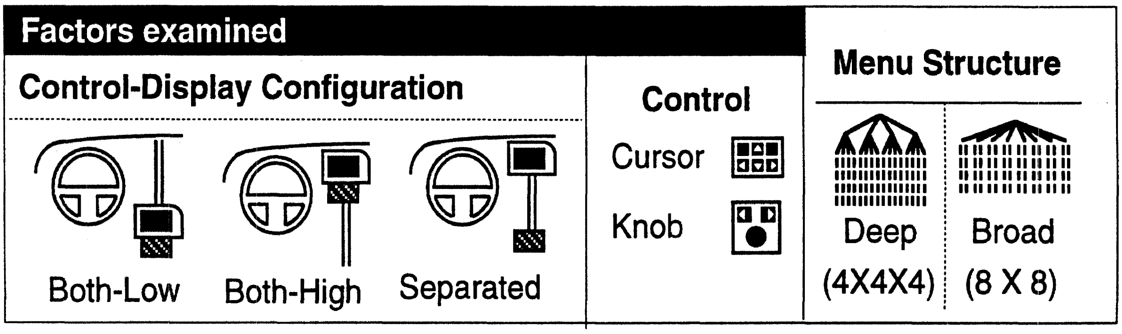
1 "... check the temperature outside the car"

2 Main Menu
Stereo
Climate...
Nav...
Vehicle...

3 Climate...
Fan
Temp
Vent
Air Filtering

4 Temp
Left...
Right...
Rear...
Outside...

5 "Beep!" (correct) - *or* - "Buzz!" (incorrect)



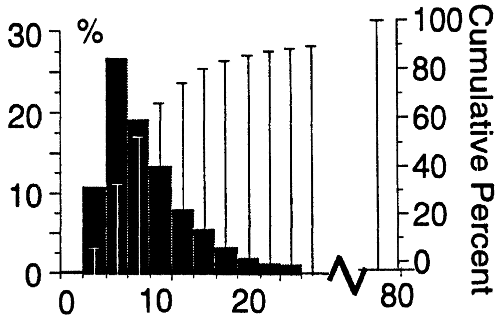
Factors examined

Control-Display Configuration
Both-Low
Both-High
Separated

Control
Cursor
Knob

Menu Structure
Deep (4X4X4)
Broad (8 X 8)

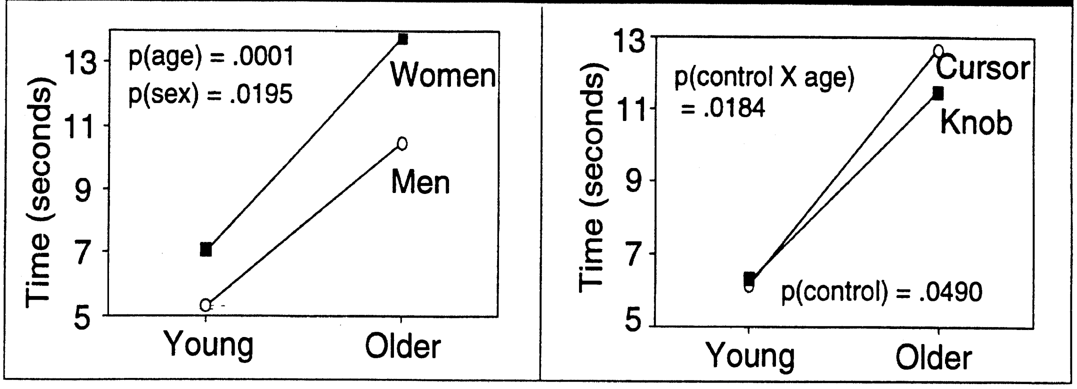
3 RESULTS



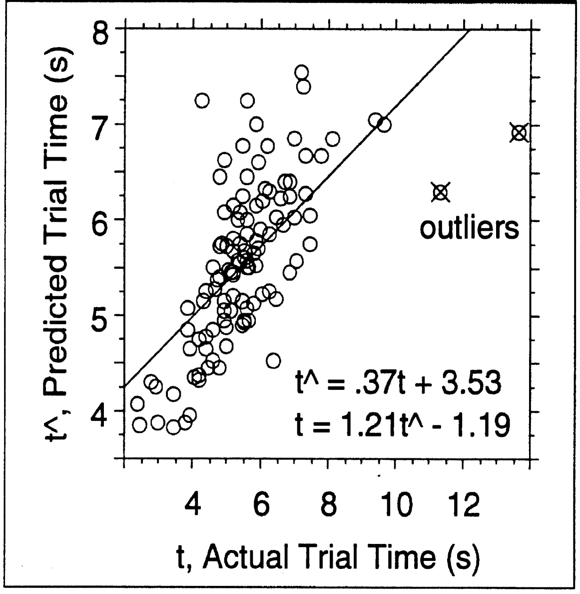
Histogram of trial times (seconds)

Mean = 9.14
Median = 7.07
Minimum = 1.27
Maximum = 87.28

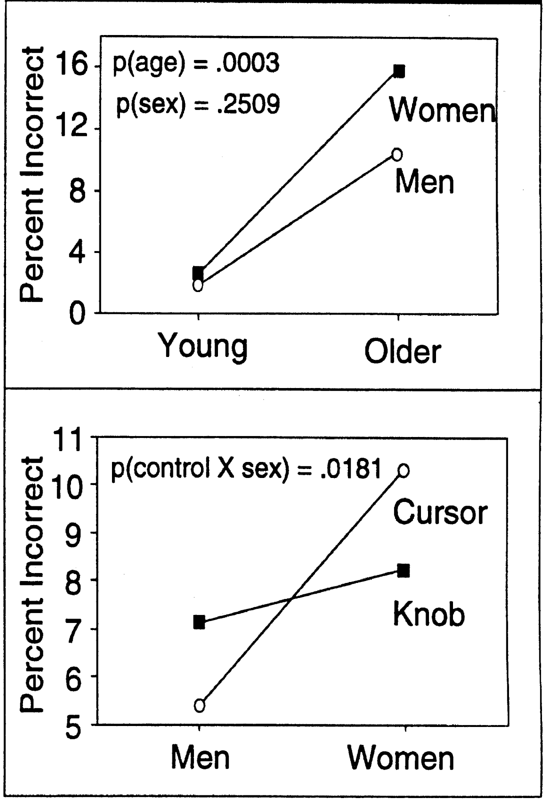
Significant effects and interactions for time



GOMS predictions vs. actual times



Significant effects and interactions for error rates



4 CONCLUSIONS

1. Trial times were almost twice as long for older drivers as for young ones (12.1 vs. 6.2 s) and were somewhat longer for women than men (10.4 vs. 7.9 s).
2. There were fewer lane excursions due to distraction for the deep menu structure than for the broad (5.00 vs. 5.67), but menu structure had little effect on time or errors.
3. The knob control yielded 9% shorter times and led to 3% fewer errors than the cursor control, both significant effects.
4. The control/display configuration had little effect on times or errors.
5. To predict actual times, multiply the GOMS estimate by 1.2 and subtract 1.2.

TABLE OF CONTENTS

INTRODUCTION....1
- Overview....1
- Literature review....2
- Issues....9

TEST PLAN....11
- Overview....11
- Test participants....11
- Test materials and equipment....11
  - In-vehicle menu interface....11
  - Menu structures....12
  - Controls....14
  - Control/display configurations....16
  - Task prompts....19
  - Driving simulator....19
- Test activities and their sequence....21

PILOT RESULTS....23
- Overview....23
- Effect of control....23
- Number of trials per block....23
- Implications....25

MAIN RESULTS....27
- Overview....27
- Menu task results....27
  - Typical selection times and error rates....29
  - Effects of age and sex....30
  - Effects of menu structure, control, and configuration....32
  - Learning effect....36
- Driving performance....39
  - Typical values....39
  - Means and ANOVA results....41
- GOMS analysis....42
- Survey results....51
- Eye fixations....54
- Initial subject reactions....54

CONCLUSIONS....57
What are typical selection times and error rates for novice users of a simulated in-vehicle menu system?....57
Can selection times be reliably predicted using a GOMS model?....57
What is the effect of driver age and sex on selection times and error rates?....57
How do selection times and error rates vary as a function of menu structure, control type, and the location of the control and display?....58
To what extent do selection times and error rates decline with practice?....58
What predictions can be made about the safety of in-vehicle menu interfaces?....59
How acceptable is the idea of an in-vehicle menu interface to younger and older drivers?....59
How many menu end nodes should be explored per block?....60
Concluding remarks....60

REFERENCES....61

APPENDIX A - SCREEN SHOT OF THE MENUPLAYER CONTROL PANEL....63

APPENDIX B - CONTROL AND DISPLAY LOCATIONS....65

APPENDIX C - TASK PROMPTS....67

APPENDIX D - PROMPT SETS....69

APPENDIX E. DRIVER INTERFACE RESEARCH SIMULATOR....71

APPENDIX F - PARTICIPANT CONSENT FORM....73

APPENDIX G - BIOGRAPHICAL FORM....75

APPENDIX H - EXPERIMENTAL DESIGN TABLE....77

APPENDIX I - ACCEPTABILITY SURVEY....79

APPENDIX J - EXPERIMENTER INSTRUCTIONS....81

APPENDIX K - ACTUAL TIMES VERSUS GOMS PREDICTIONS....85

INTRODUCTION

Overview

There is growing interest in organizing the driver interface for secondary functions (e.g., climate control subsystems, navigation, and entertainment) as a collection of hierarchical menus. Although the menu concept has been around for some time (Green, 1979), only recently has the number of in-vehicle functions grown to the point where there is insufficient instrument panel space for dedicated controls and displays. For many of the features being considered, there is either strong customer demand or there are major opportunities to enhance vehicle safety (Green, Serafin, Williams, and Paelke, 1991; Green, 1996).

Several systems on the market already utilize hierarchical menu architectures. These include navigation systems (sold primarily in Japan) and several systems that integrate entertainment and climate control units. In the United States, one of the best known prototypes of this architecture is the Delco Eyes-Forward interface (Heuchert, 1995), an attempt to integrate audio, climate, navigation, and trip controls into a single, menu-based interface (see Figure 1). It was designed to be installed in the speedometer/tachometer cluster and was operated via two pairs of rocker switches on each side of the steering wheel.

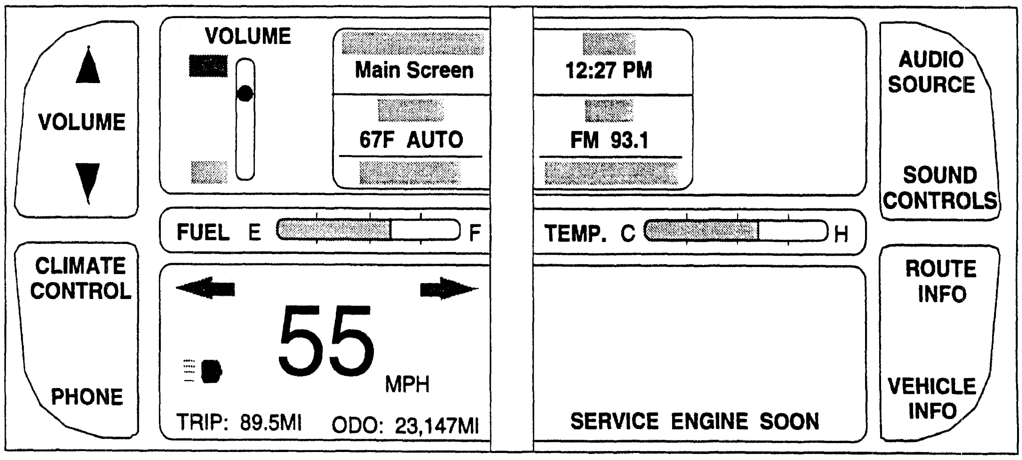


Figure 1. The Delco Eyes Forward Interface.

The safety and usability of in-vehicle menu systems, however, is very much in question. Although menu-based interfaces have many desirable qualities, several criticisms have been leveled against them. These include the following:

1. Hierarchical menus will be too complicated for drivers to learn, even when they devote their full attention to the interface.
2. Because of their complexity, only younger drivers with computer experience may understand the basic concept.
3. Even if drivers can operate menu-based interfaces, the interfaces will require so much attention that they might be unsafe to operate while the vehicle is in motion, distracting drivers from the primary tasks of steering, speed maintenance, and avoiding hazards.

4. Market demand will be low because the interface style is so different from that which drivers are familiar.

These claims have not been well tested in the literature. Since the performance of a test system would be as much a reflection of that particular system as menu-based interfaces in general, it is crucial that the test system be as well designed as possible. Hence, before any claims can be properly investigated, design principles and recommendations for menu interfaces are sorely needed. The development of such principles is a major goal of this project.

Literature review

Although the literature directly relevant to the use of computer menus while driving is limited, there is a considerable body of literature on menu design in general. The best summary of the menu literature is in The Psychology of Menu Design (Norman, 1991). Primary design issues are shown in Table 1.

One critical issue that has been well explored in the literature is menu breadth (the number of items at each level) versus depth (the number of levels of menus). Specifically, suppose one had 64 end nodes in the menu hierarchy. Would the best design be to have one menu of 64 items (1 x 64), two levels of menus with eight items (8 x 8) at each level, three levels with four items each (4 x 4 x 4), or six levels with two items each (2 x 2 x 2 x 2 x 2 x 2)? The decision would depend upon the particular logical groupings of menu items, the space available for each menu, the type and location of the control used to select items, and the time available for users to read through the menu.

Table 1. Issues of menu design.

Issue	Options
How should menus be called?	Pull down, drop down, pop up
If pull down/drop down, how should they be organized?	Linear, circular
If pull down/drop down, where should the menu bar be located?	Screen border, window border
How should menus disappear?	Release mouse button, click mouse button, etc.
Which options should be shown?	All, all available, all but gray out those not available
How should options be shown?	Alphabetic, categorically (possibly with block clustering), frequency of use, importance of use, random, etc.
How should options be selected?	Mouse click, mouse drag and release, touch screen, enter entry number, cursor down and press enter key, etc.
Should multiple selections be allowed for last menu?	No (allow only one), select range via dragging, discontinuous
Should menus show single or multiple sets of options?	Single, multiple
Should submenus be allowed?	Yes, no
If allowed, how should submenus be shown?	Fan out—top aligned with current choice, middle aligned with current choice, etc.
How should menu cells be sized?	All equal height, Fittsized (increasing height)
How should the menu system be structured for navigation?	Hierarchical, connected
How should menus and menu entries be named?	Single word title, multiple word title, name + 1 example, name + multiple examples
Should keyboard shortcuts be allowed?	For all entries, for some entries, for no entries
What should the shortcuts be for?	Menu entries, terminal items

Miller (1980, 1981), in an attempt to address this issue, had subjects find items in a series of unordered lists ranging from two to 64 items. He proposed that selecting an item from a menu is a linear function of the number of choices on screen at any given time. Further, he suggested that selecting items typically involves one or more category matches (decisions about which category the goal item belongs to) followed by an identity match (a decision about which menu item is equivalent to the goal item). For example, a biologist looking up information on monkeys might need to select "Animals," followed by "Mammals," and finally "Monkeys" from a three-level menu system (two category matches and an identity match). To estimate the time of such a task, Miller proposed the following two formulas:

$CM = .80 + .035N$
$IM = .32 + .11N$

where,

CM = the time to perform a category match (in seconds)
IM = the time to perform an identity match (in seconds)
N = the number of choices for the particular menu

In related work, Landauer and Nachbar (1985) postulated that the time to select information from a touch screen (in an office-like setting) could be predicted from a combination of Hick's Law (also referred to as the Hick-Hyman Law) and Fitts Law. Hick's Law (Hyman, 1953) states that the time to select an item from a collection of items (choice response time) is as follows:

$RT = c + k * \log_2 x$

where,

RT = response time (usually measured in seconds or milliseconds)
x = number of equally likely alternatives
c, k = empirically determined constants (k decreases with practice)

When alternatives are not equally likely, their probability must be taken into account. In such cases, choice response time is expressed using equations based on information theory, typically as follows:

$RT = a + b * H$

where,

H = information (bits) = $\Sigma_i (p_i * \log_2(1 / p_i))$
p_i = probability of item i
a, b = empirically determined constants

Note: $\log_2 a$ can be converted to any base using $\log_b a = \log_x a / \log_x b$

Thus, the key points from Hick's Law is that decision time is a log function of the number of alternatives, with the parameters of the prediction being empirically determined for each context. Fitts Law (Fitts, 1954), which has a similar form, is extremely robust and works for all limbs in a variety of environments (including underwater and zero gravity), and even works when remote manipulators are used. Fitts Law states that the time to move a limb towards a target can be determined by the following expression:

$$MT = c_1 + c_2 * \log_2(2d / w)$$

where,

d = movement distance
w = target width plus allowable movement error (same units as d)
c_1, c_2 = empirically determined constants

In the Landauer and Nachbar experiment, two sets of goal stimuli were used: (1) integers from one to 4096 and (2) 4096 words from four to 14 characters long. On each trial, the goal item was presented followed by a screen on which either 2, 4, 8, or 16 items appeared. Except for the last screen (showing only single items) all other screens indicated ranges of choices (one to 1024 for digits, e to h for letters). Subjects touched the place on the screen containing the appropriate response. This experiment yielded four equations (see Table 2). All expressions were clearly linear as a function of the log of the number of choices (menu items on screen) since the menus were well structured (ordered), not the case for Miller's menus.

Table 2. Landauer and Nachbar's equations of menu selection time.

Stimuli	Session	Time per choice (ms)
Words	1	1338 + 826X
	2	1177 + 629X
Numbers	1	820 + 575X
	2	711 + 517X

X = log(number of alternatives on screen)

By way of comparison with Miller's data, Landauer and Nachbar predict a mean selection times of 5.00 seconds for the hierarchy for Session 1 and 2.48 seconds for Session 2. This second prediction is quite close to the value offered by Miller (2.75 seconds). The classification task in Miller's experiment was more complex than that of Landauer and Nachbar. For example, one of the trials in Miller's experiment required the subject to select the menu item that would contain information on Plato (the choices included agriculture, medicine, physics, zoology, country, topography, art, and person). There were also some minor differences in timing and the response method (buttons for Miller's experiment and a touchscreen for Landauer and Nachbar's). These differences emphasize the need to collect data for each context.

Two key experiments concerning menu design were conducted by Card (1981, 1984). Subjects were asked to search menus organized randomly, functionally (by category), or alphabetically. The results established that search was random with replacement (unsystematic). In other words, each search instance involved choosing among all menu items, not just those that had not been examined in previous instances. Hence, the probability of finding the target after time T was:

$$P(T) = 1 - e^{-m(T - t_0)}$$

where,

m = $-\ln(1- p_1)\,\tau$
t_0 = nonvisual part of the search task (pressing a button)
p_1 = probability of finding the target on the first fixation
τ = mean time per fixation including transitions.

For the example subject described by Card, τ was approximately 260 ms per saccade (eye fixation), t_0 was 1.6 seconds, and p_1 was 0.26.

On average, search times were least for alphabetically grouped lists, greater for functionally grouped lists, and most for random lists. Differences in search times between organization schemes diminished with practice, however. Card also examined grouping effects within menus and found recall differences between and within groups (chunks). His data highlights the value of modeling eye fixation patterns, models particularly important to driving.

Musseler (1994) describes two experiments in which subjects used keyboard shortcuts to retrieve items from a drop down menu. Although the research presented in this report did not focus on semantic issues, Musseler's work also concerns menu structure as well. Figure 2 shows the English translation of the German menu items. Tables 3 and 4 list the various model factors, their descriptions, and coefficients for predicting selection times.

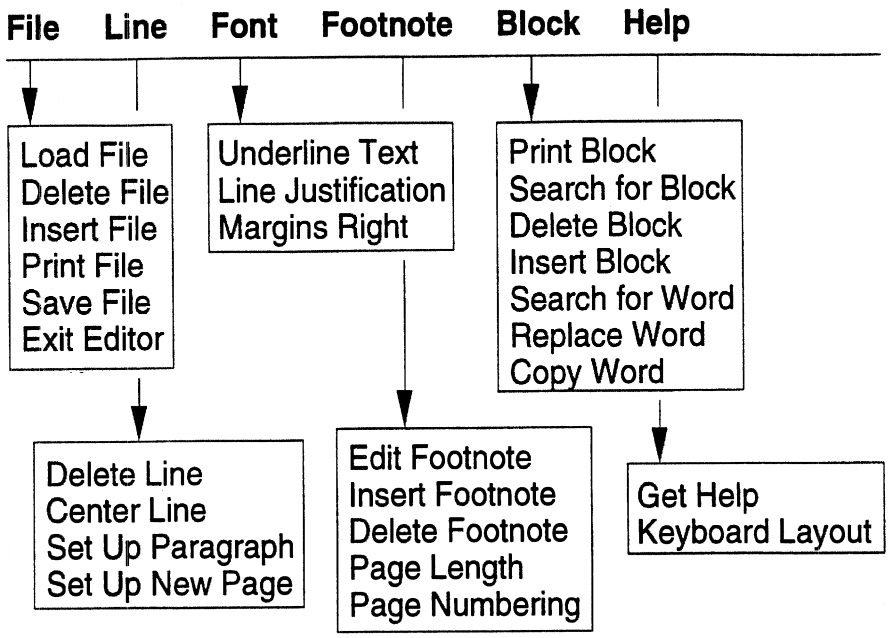


Figure 2. Menu structure used by Musseler.

Table 3. Factors examined by Musseler.

Factor	Symbol	Comment
Position in menu bar	PM	6 positions
Position in submenu	PS	1-6
Number of submenu items	NS	2-6 items
Fanning	AF	Action fanning—verb repetition (e.g., delete file, delete footnote)
	OF	Object fanning—noun repetition (e.g., load file, delete file)
Overtness	OC	Overt (load file in File menu) versus covert (exit editor in File menu)
Subgrouping	SMG	2 subgroups (block and word) in block menu
Menu item length	LMB	4-8 letters
Submenu item length	IL	9-19 letters
Longest item in submenu	WSM	11-19 letters

Table 4. Experimental coefficients (in ms) for predicting selection times.

	Experiment 1		Experiment 2	
Factor	Menu Select	Submenu Select	Menu Select	Submenu Select
PM	68	1	78	7
PS	9	125	-18	-73
NS	126	-25	-242	197
AF	123	33	-30	-104
OF	115	299	812	115
OC	426	-101	426	177
SMG	595	-136	-1524	515
IL	93	15	99	21

Note that fanning, overtness, and subgrouping typically had larger effects (coefficients) than commonly explored factors such as the number of items to choose from and the name length. Assuming that (1) the items of interest are in the middle of the list, (2) no differences accrue when a second level submenu is used, (3) the longest menu item has eight characters, and (4) menus have neither action nor object fanning, the estimated selection time is 2.92 seconds using the equation for experiment 1, 2.72 seconds for experiment 2, reasonably close to the expressions from the other studies discussed here.

For studies of human-computer interfaces, the most popular computational approaches involve the Model Human Processor and Keystroke-Level Models (Card, Moran, and Newell, 1983). The Model Human Processor assumes human performance can be modeled using three processors (perceptual, cognitive, and motor), each with their own cycle times, and four memory systems (visual and auditory sensory stores, short-term memory, and long-term memory), each with their own decay constants, storage capacities and codes, and retrieval mechanisms. For many tasks, the Model Human Processor analysis is too fine grained. In the Keystroke Model, the primary elements are thinking (mental preparation), keying, and pointing. The model parameters are specified in Table 5.

Although these models are well described, applying them to driving is a challenge. Both of these models were developed to examine single-task performance, but when driving is also involved, the dual task context needs to be considered. Further, secondary control actions in motor vehicles rarely involve touch typing on a keyboard or using a mouse, controls commonly used for interacting with a computer. Nonetheless, even if many of the basic model assumptions are violated, especially the single task assumption, the rank order of model predictions should agree with the actual values measured.

Table 5. 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

In fact, Paelke (1993) and Paelke and Green (1993) reported reasonably good agreement between the rank order predictions (from Keystroke-Level models) of the time to enter destinations into a navigation system and the actual time required by drivers. Good predictions were also reported by Manes, Green, and Hunter (1997) for the Siemens Ali-Scout navigation system.

Finally, in a study very closely related to this project, Sumie, Li, and Green (1997) conducted three experiments to examine how a hierarchical menu system should be tested. The first two experiments concerned matching driving performance in a simulator with performance on the road and examining learning in the simulator. In the third experiment, 16 subjects retrieved information from a hierarchical menu system while driving. The information entered was entirely numeric, however, which is not typical of real menu items. Also, real vehicle features were not simulated. Issues examined included the control type (knob, trackball, touchpad, and number pad) and the menu structure (two levels of eight items, three levels of four items, six levels of two items). Both driving performance measures and keying times were recorded. Analysis of that data is in progress.

Issues

The following issues were selected based on the gaps in the literature (especially concerning menu use in a timesharing or driving context), the desire to build engineering models of driver performance, the need to learn how menu systems can be optimized prior to comparing them with other architectures, and the specific interests of the sponsor:

1. What are typical selection times and error rates for novice users of a simulated in-vehicle menu system and can times be reliably predicted using a GOMS model?
2. How do selection times and error rates vary as a function of menu structure, control type, and the location of the control and display?
3. What is the effect of driver age and sex on selection times and error rates?

4. To what extent do selection times and error rates decline with practice?
5. What predictions can be made about the safety of in-vehicle menu interfaces?
6. How acceptable is the idea of an in-vehicle menu interface to younger and older drivers?

These issues were investigated in two experiments. The first, a pilot experiment, was conducted to help make decisions about a subsequent main experiment, namely, which controls to use and how many trials per block to include. For both, subjects performed tasks on a simulated in-vehicle menu interface while operating a driving simulator.

TEST PLAN

Overview

The experimental protocol for both the pilot and main studies is discussed in this section. The design of the two studies were similar. The differences included the number of participants, the control types tested, and the number of trials used per block. These differences are fully described below.

Test participants

Three young men (ages 22, 22, and 26) and one young woman (age 23) participated in the pilot experiment. All participants were licensed drivers and all had regular access to a vehicle. Annual mileage driven was between 6,000 and 10,000 (mean of 9,000), and years of driving experience ranged from 6 to 10 (mean of 7). Weekly computer usage ranged from 25 to 40 hours (mean of 31). Finally, corrected visual acuity ranged from 20/13 to 20/17 (mean of 20/15). The subjects were all UMTRI employees with prior simulator experience but none were associated with this project. Each was paid his or her regular hourly wage for about 2.5 hours of testing.

There were 24 participants in the main experiment, 12 young (30 and under) and 12 older (65 and above). The young participants ranged from 21 to 27 years old (mean of 23) while the older particpants ranged from 65 to 70 (mean of 67). Within each age group, there were an equal number of men and women. All participants were licensed drivers and all but one had regular access to a vehicle. Annual mileage was between 4,000 and 30,000 (mean of 13,400), and years of driving experience ranged from three to 52 (mean of 27). Weekly computer usage ranged from 0 to 70 hours (mean of 17). Finally, corrected visual acuity ranged from 20/13 to 20/70 (mean of 20/29). The subjects included both new recruits and those who had served in previous UMTRI studies (with seven of 24 having driven the simulator before). Each was paid $30 for between 1.5 and 2.5 hours of testing.

Test materials and equipment

In-vehicle menu interface

The simulated in-vehicle menu interface (called MenuPlayer) was written using Allegiant SuperCard for the Macintosh. Three input devices (a cursor control, number pad, and knob control) were used to select entries from menus shown on a Cascade Technologies DiscoveryMATE R65 six-inch LCD display (provided by the sponsor). The display was mounted to a custom stand.

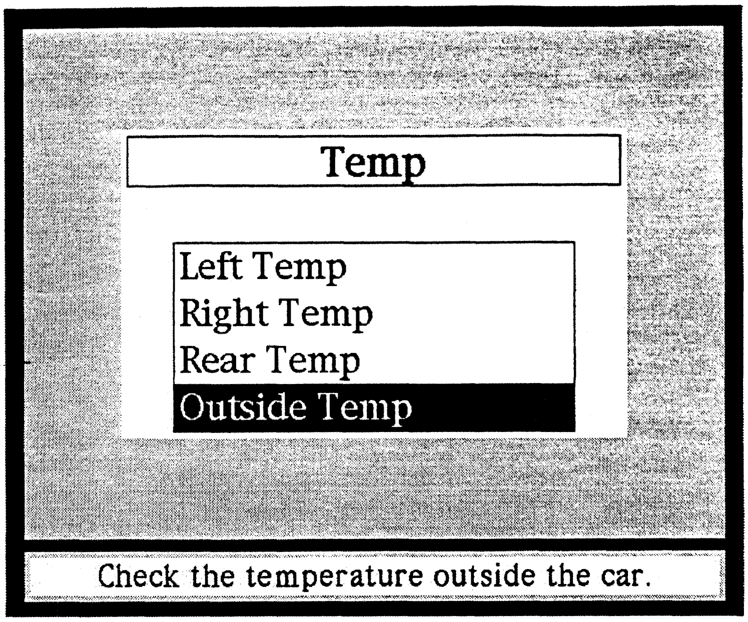
The simulation (1) presented the target item both auditorally and visually, (2) displayed menus, (3) processed input from the driver, and (4) collected data for each movement of the control. For each trial, MenuPlayer would chime and play a short clip of digitized speech asking the driver to perform a specific task (e.g., “How would you eject the tape?”). In case the driver did not hear the entire spoken request, the associated text also appeared at the bottom of every menu. Then, using the knob or cursor control, the driver would select menu items until the task was completed. Finally, the program either sounded a confirming "beep" for correct selections or a "buzz" (from a television game show) for incorrect ones. After a five-second delay, the next trial would begin automatically. (See Appendix A for a screen capture of the experimenter's interface.)

Menu structures

The same two menu structures were used in both the pilot and main experiments (see Tables 6 and 7 and Figures 3 and 4). These included a deep structure with four items per menu and three levels (4 X 4 X 4) and a broad structure with eight items per menu and two levels (8 X 8). Hence, the deep structure required three menu selections to complete a given task while the broad structure required two.

Table 6. The deep (4 X 4 X 4) menu structure.

Main Menu	Submenu	Final Menu
Stereo	Volume/Tone	Volume, Bass, Treble, Balance
	Radio	Tune, Seek, Preset, Scan
	CD	Skip Track, Scan Tracks, Skip Disk, Random Mode
	Tape	Fast Forward, Rewind, Tape Direction, Eject Tape
Climate Control	Fan	Fan Off, Auto, Econ, Fan Speed
	Temp	Left Temp, Right Temp, Rear Temp, Outside Temp
	Vent	Panel, Floor, Panel & Floor, Defrost
	Air Filtering	Fresh Air, Recirculate, Mixed Air, Filtered Air
Navigation	Map Settings	Zoom In/Out, North Up, Heading Up, Show Names
	Set Destination	Preset Location, Enter Address, Find Business, Locate on Map
	Route Options	Shortest Route, Fastest Route, No Highways, Scenic Route
	Alert Sound	Voice Alert, Chime Alert, Voice & Chime, No Alert Sound
Vehicle Setup	4WD Mode	Full-Time 4WD, 4WD High, 4WD Low, 2WD
	Shift Mode	Economy, Normal, Power, Hold Gear
	Steering	Easy, Light, Medium, Firm
	Ride	Normal, Touring, Sport, Off Road

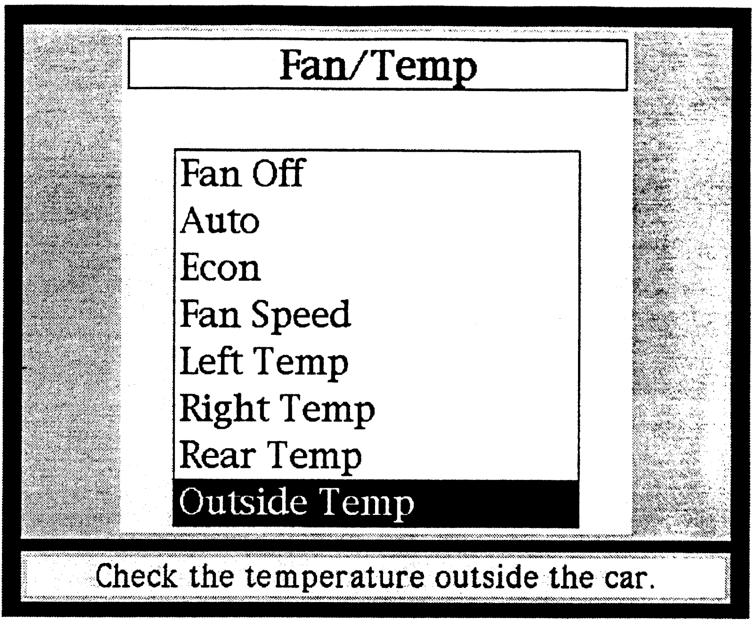


Check the temperature outside the car.

Figure 3. The Temp menu of the deep menu structure with "Outside Temp" selected (as seen by subjects).

Table 7. The broad (8 X 8) menu structure.

Main Menu	Final Menu
Volume/Radio	Volume, Bass, Treble, Balance, Tune, Seek, Preset, Scan
CD/Tape	Skip Track, Scan, Tracks, Skip Disk, Random Mode, Fast Forward, Rewind, Tape Direction, Eject Tape
Fan/Temp	Fan Off, Auto, Econ, Fan Speed, Left Temp, Right Temp, Rear Temp, Outside Temp
Vent/Air Filtering	Panel, Floor, Panel & Floor, Defrost, Fresh Air, Recirculate, Mixed Air, Filtered Air
Map/Set Destination	Zoom In/Out, North Up, Heading Up, Show Names, Preset Location, Enter Address, Find Business, Locate on Map
Route/Alert Sound	Shortest Route, Fastest Route, No Highways, Scenic Route, Voice Alert, Chime Alert, Voice & Chime, No Alert Sound
4WD/Shift Mode	Full-Time 4WD, 4WD High, 4WD Low, 2WD, Economy Shift, Normal Shift, Power Shift, Hold Gear
Steering/Ride	Easy Steering, Light Steering, Med Steering, Firm Steering, Normal Ride, Touring Ride, Sport Ride, Off Road



Check the temperature outside the car.

Figure 4. The Fan/Temp menu of the broad menu structure with "Outside Temp" selected (as seen by subjects).

For both structures, there were a total of 64 final items (end nodes), and they were the same in each case, except for 10 items in the last three menus of the broad structure where a clarifying word was added. The deep menu structure consisted of a main menu with broad categories, four submenus with subcategories, and 16 final menus with the end nodes. The broad menu structure consisted of a main menu that was constructed using pairs of items from the submenus of the deep structure (separated by a slash) and eight final menus.

The menus were designed to resemble those of real in-vehicle interfaces, such as the Delco Eyes Forward (Heuchert, 1995), the Siemens Quickscout, and the Rockwell PathMaster. However, several features that would likely be given dedicated controls in a real interface (e.g., volume and temperature) were included in the menus to allow for experimental counterbalancing. (Since the purpose of the experiment was to cleanly evaluate experimental factors, not to test a real product, this compromise seemed appropriate.) For the same reason, some of the items were somewhat unconventional (such as the "off-road" ride and "filtered air" settings), but they were generally found to be understandable and fit well into the two hierarchies. Finally, menu item names were chosen to be as short and nontechnical as possible to minimize reading time and confusion.

Controls

The menus were navigated using three different controls—a cursor control, a number pad, and a knob control. Because cost and schedule constraints did not permit all three to be examined in the main experiment, the cursor control and number pad were tested in the pilot experiment (see Figure 5), and the better performing of these two (the cursor control) was compared with the knob control in the main experiment (see Figure 6). The comparison of the cursor control with the number pad was similar to that explored by Shinar, Stern, Dubis, and Ingram (1985), except that their direct-selection method was alphabetic (first character of each word) instead of numeric.

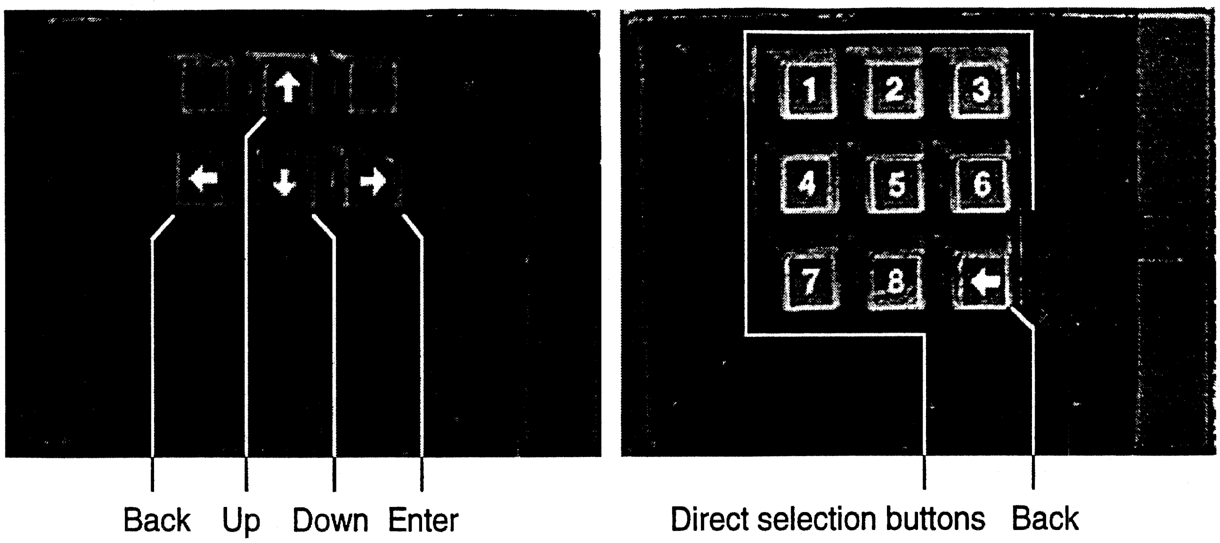


Figure 5. The controls tested in the pilot experiment—the cursor control (left) and the number pad (right).

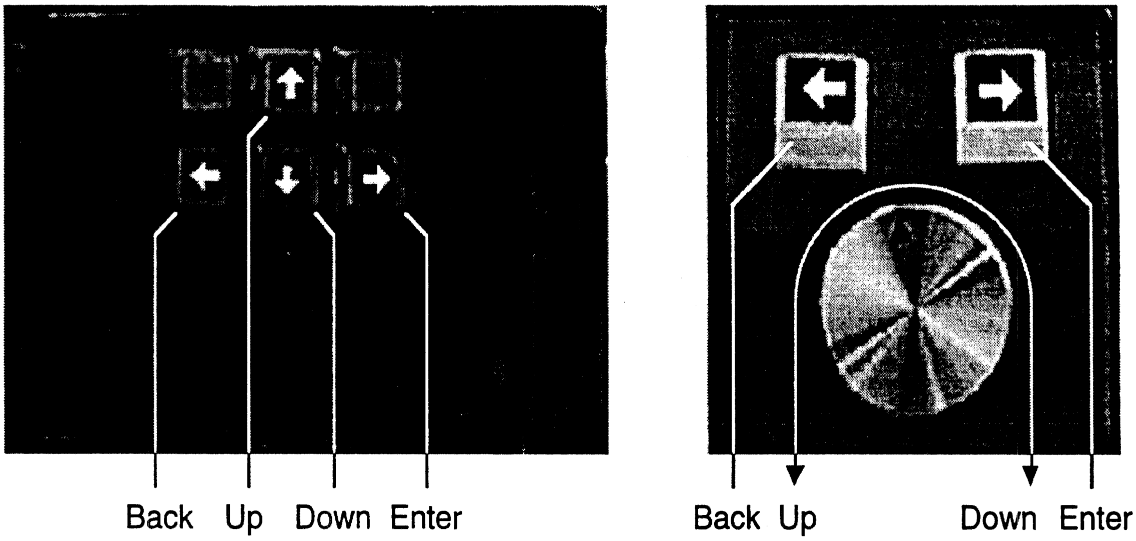


Figure 6. The controls tested in the main experiment—the cursor control (left) and the knob control (right).

Using the cursor control involved pushing the up and down arrows until the black selection bar (see Figures 3 and 4) was over the desired item, then pressing the right arrow (enter button) to select it. The number pad, on the other hand, involved simply pressing the button corresponding to the number of the desired item. When the number pad was used, each menu item was numbered so that the driver would not have to count down the list. Finally, the knob control functioned similarly to the cursor control, except the selection bar was moved up and down using the knob instead of scroll keys. While pushing the knob to make selections (as with the Siemens Quickscout) would have been preferable, it was not possible to fabricate such a switch in the project time frame. For all three controls, pressing the left arrow (back button) brought up the previous menu.

To minimize complexity and ease analysis of the keystroke data, the selection bar did not wrap around (jump from the last item to the first or from the first to the last) for either control. Also, arrows were used on the buttons instead of names (such as "Enter" and "Back") because they fit well into the space available and were discriminable from each other. Any problems understanding them were overcome with limited practice, and subjects rarely looked at the labels once the main experiment began. In a production system, other labels might be chosen.

The cursor control was a Sophisticated Circuits POWERPad (a number pad) with all but six of the key caps removed. Black tape was used to hide the exposed sockets, and custom labels were placed over the remaining caps. The number pad was constructed in essentially the same way. The knob control was custom made using a metal case, a 12-position rotary switch with detents, and two single-push, single-throw, momentary buttons. Because these buttons provided minimal feedback, MenuPlayer was programmed to make a "clicking" sound each time one of them was pressed.

Control/display configurations

To explore the effects of control and display location and their relationship to each other, three different control/display configurations were used in both the main and pilot experiments (see Table 8). These were (1) a *both-high* configuration where the display was mounted 5.0 inches above the center of the steering wheel and the control was mounted just underneath it (see Figure 7), (2) a *both-low* configuration where the display was mounted 3.0 inches below the center of the steering wheel and the control was placed towards the front of the arm rest (see Figure 8), and (3) a *separated* configuration where the display was placed as in the both-high configuration and the control as in the both-low (see Figure 9). (See Appendix B for drawings showing the control and display locations.)

Table 8. Positive and negative aspects of the control/display configurations.

	Control High	Control Low
Display High	Both-high configuration • Shorter eye movements between road and control/display • Good control/display association • Nonideal control location	Separated configuration • Shorter eye movements between road and display (but not road and control) • Poor control/display association • Good control location
Display Low	(Not tested) • Arms could block view of display • Longer eye movements between road and display • Poor control/display association • Nonideal control location	Both-low configuration • Longer eye movements between road and control/display • Good control/display association • Good control location

A mounting bracket was glued to the back of the cursor control so that it would stand up at a 30 degree angle when placed on the arm rest and so that it could be mounted to the display. Velcro was used to mount the controls to the bottom of the display for the both-high configuration. For the both-low configuration, a large binder clip was used to prevent the cursor control from sliding forward on the arm rest. Similarly, a C-clamp, an angled block of wood, and Velcro were used to hold the knob control in place. For the both-high configuration, an additional armrest was placed on top of the main armrest to minimize fatigue caused by the high location of the control.

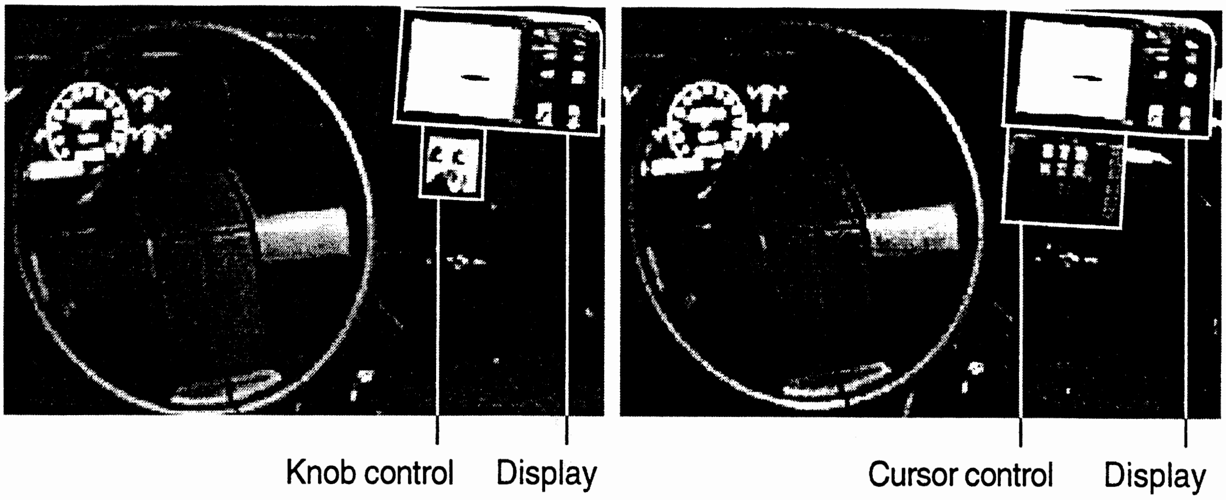


Figure 7. The both-high configuration with each control.

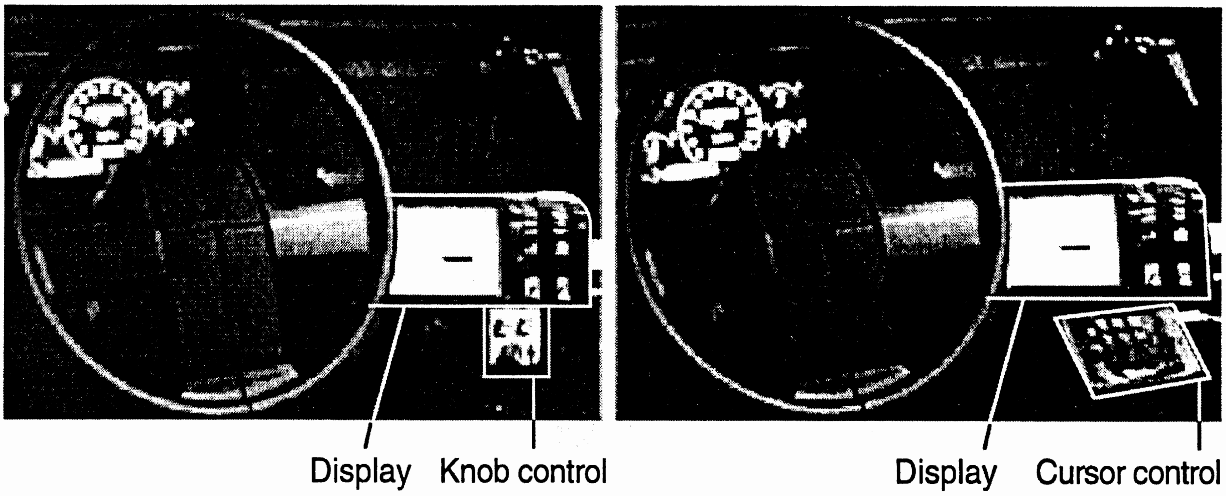


Figure 8. The both-low configuration with each control.

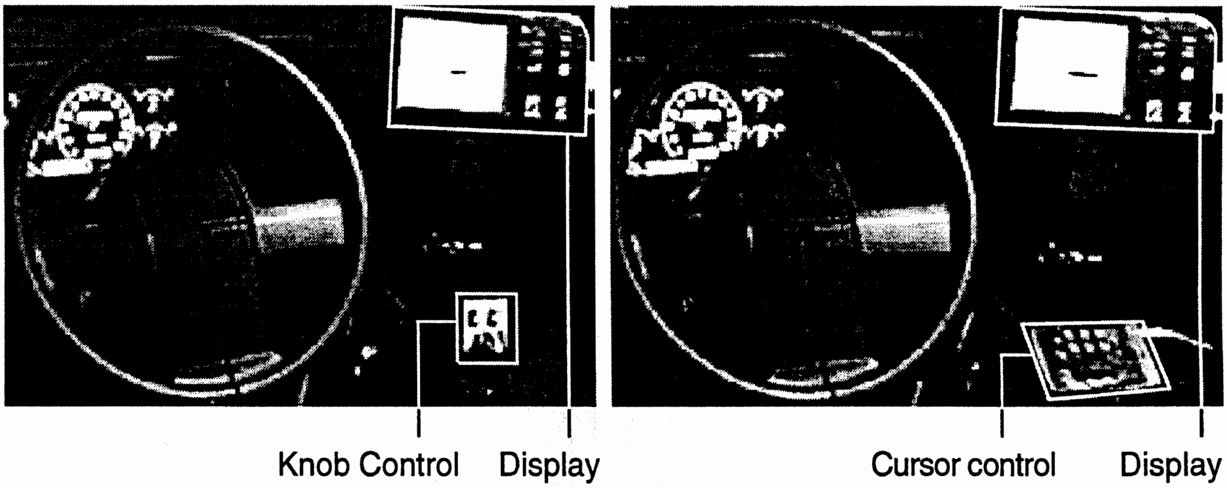


Figure 9. The separated configuration with each control.

Task prompts

The prompts consisted of 64 sound clips (about two to five seconds each) recorded with a male voice, one for each end node of the menu hierarchy (see Appendix C). Each sound clip was a request for the driver to perform a task (e.g., "How would you check the temperature outside the car?"). The wording was chosen to be as nontechnical and natural as possible and to provide enough cues so that the driver could find items in the menu structure even if they were unfamiliar with the item. Sound clips were played at the beginning of the trial, and in case the driver did not hear the message properly or forgot it, a reminder message was provided at the bottom of the display (see Figures 3 and 4).

Six prompt sets (random sequences of sound clips) were created for each of the two studies. For the pilot experiment, each prompt set consisted of 32 items (see Table 23 in Appendix D). For these, participants encountered two items from each of the 16 final menus of the deep structure (equivalent to four items from each of the final eight menus of the broad structure). Time constraints prevented the use of 64-item sets, which would have allowed exploration of all menu end nodes.

Because analyses of the pilot data revealed that 16-item prompt sets would probably have been as effective as the 32-item sets (and would have allowed for much shorter session lengths and less driver fatigue), 16-item sets were used in the main experiment (see Table 23 in Appendix D). Accordingly, one item for each of the 16 final menus of the deep structure was explored (equivalent to two items from each of the final eight menus of the broad structure).

In both experiments, each prompt set appeared twice since there were 12 blocks total and one prompt set was presented each block. In addition to the primary prompt sets, two randomized, 32-item practice sets were also used, each appearing once per subject. One contained the even-numbered prompts and the other contained the odd.

Driving simulator

This experiment was conducted in the UMTRI Driver Interface Research Simulator, a low-cost driving simulator based on a network of Macintosh computers (MacAdam, Green, and Reed, 1993; Green and Olson, 1997; Olson and Green, 1997). The simulator (see Figure 42 in Appendix E) consisted of an A-to-B pillar mockup of a car, a projection screen, a torque motor connected to the steering wheel, a sound system to provide engine, drive train, tire, and wind noise (see Figure 43 in Appendix E), a computer system to project images of an instrument panel, a vibration system to simulate road feel, and other hardware. The projection screen, offering a 30 degree field of view, was 20 feet (7.3 meters) in front of the driver, effectively at optical infinity. Finally, driving performance measurements (including lane position, speed, and heading) were logged at 30 hertz.

The video rack (see Figure 44 in Appendix E) and cameras (see Items 16 and 17 of Figure 42 in Appendix E) allowed for a quad-split video image to be recorded (see

Figure 10). These simultaneously showed the display, the road, the driver's face, and the instrument panel.

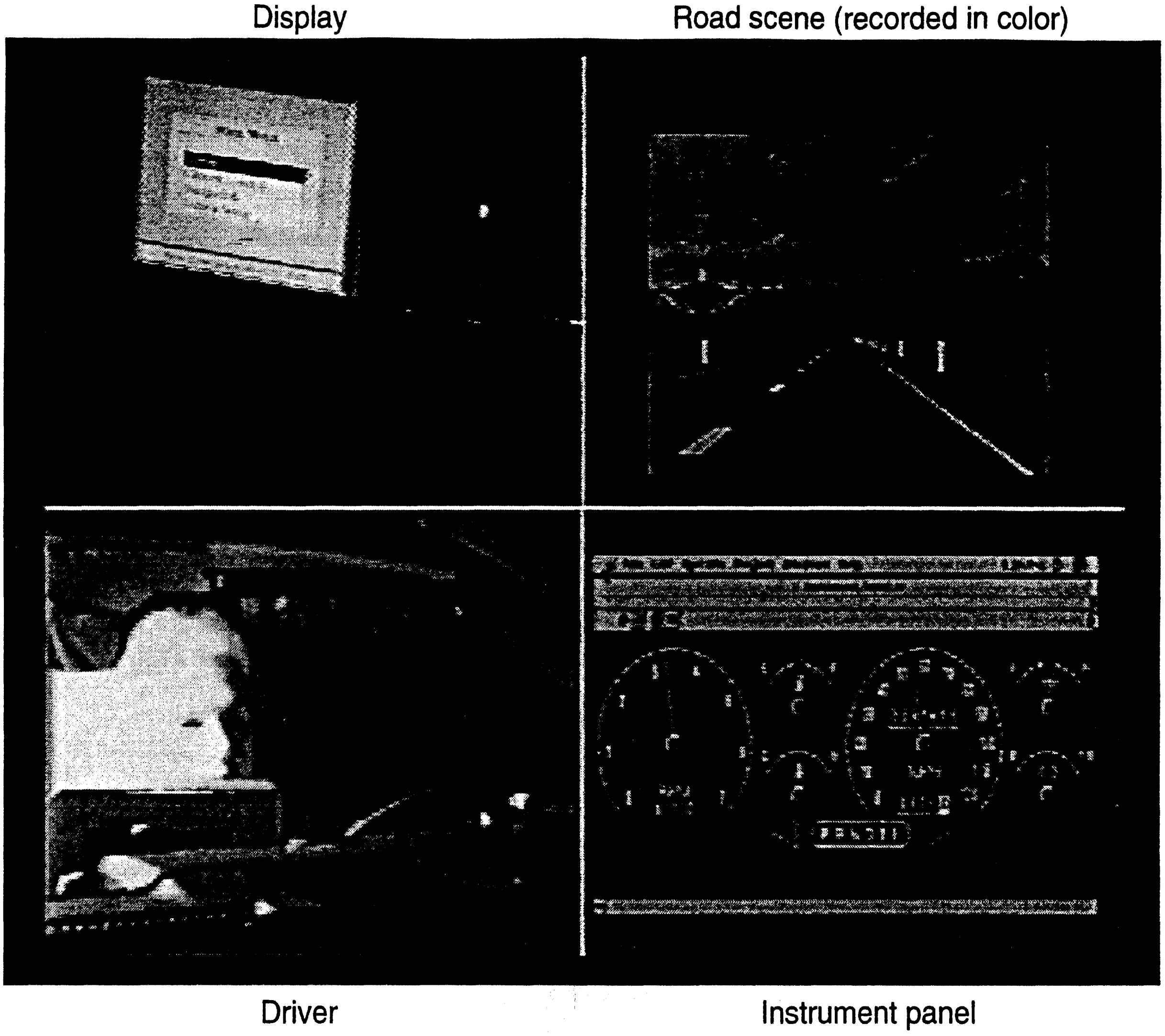


Figure 10. Quad image showing a driver glancing at the display.

The road used in this experiment (see Figure 11) was designed to be driven at 55 miles per hour. It consisted of straight sections and randomly placed large-radius curves. White posts, speed limit and route signs, and oncoming cars were placed on or at the side of the road to provide a realistic context and discourage drivers from wandering out of their lane. These objects were placed on an occasional basis except the posts which were placed every 30 feet.

Figure 11. Typical road scene showing an oncoming car and white posts.

Test activities and their sequence

The same sequence of steps were followed for both the pilot and main experiments. After filling out consent and biographical forms (see Appendices F and G), the participant was given a far visual acuity test using Landolt Rings. Next, the subject was directed to the driving simulator where he or she drove for two minutes to get accustomed to the simulator dynamics. Then, with the simulator turned off, the experimenter briefly explained the purpose of the menu interface, taught the subject how to use each control, and demonstrated the two different types of menus.

The participant next performed two 32-trial practice blocks while driving the simulator. This allowed him or her to experience all 64 end nodes, both controls and menu structures, and the both-low and both-high configurations. If subjects had initial difficulty understanding how to operate the system, the experimenter provided step-by-step instructions for each trial until the participant could continue on his or her own. Normally, this did not involve more than the first five or six trials of the first practice block.

Both experiments involved 12 blocks of trials, but the pilot experiment used 32 trials per block while the main used 16. The blocks were counterbalanced such that each participant was exposed to each combination of menu structure, control type, and configuration once and each prompt set twice. (See Appendix H for the experimental design table.) The subject was instructed to maintain a speed of 55 miles per hour, stay in the right-hand lane at all times, and make controlling the car his or her highest priority. To gain a sense of the glancing behavior of the drivers, the experimenter observed approximately four randomly selected trials for each block and recorded the minimum and maximum number of times the subject glanced at the display per menu selection.

Following the experiment, the participant filled out a survey (see Appendix I) to evaluate the menu interface and rank different alternatives for controlling in-vehicle systems. (See Appendix J for detailed experimenter instructions.)

PILOT RESULTS

Overview

The pilot experiment was conducted to make two decisions about the design of the main study—(1) which control to test against the knob control (the cursor control or number pad) and (2) how many trials per block to use (16 or 32). The first decision was made based on trial times and an ANOVA. The second involved performing several analyses to determine if half the data (16 trials) would yield the same conclusions as all the data (32 trials).

Effect of control

A repeated-measures ANOVA of time was performed whose model contained menu structure, control, configuration, and all interactions between them. With all available data included in the analysis, trial times were significantly shorter for the cursor control than for the number pad—4.18 and 4.79 seconds, respectively (p=0.044). Error rates were not analyzed because there were so few errors across the four subjects.

Number of trials per block

To answer the question of whether 16 trials per block (rather than 32) would be sufficient to obtain stable results, data analyses performed using all 32 trials were compared with analyses using only 16 of the 32 trials. Because the 32-item prompt sets used in the pilot experiment (see Table 24 in Appendix D) contained a pair of items from each final menu of the deep structure, the 16 trials were selected by choosing one of the two items from each pair at random. Further analyses were conducted for a second half of the data—the complement of the first one. For convenience, these two subsets of data are referred to as Half 1 and Half 2, respectively.

The first comparison involved plotting histograms of trial times (see Figures 12, 13, and 14) in order to visually compare the distributions for all the data with those for Half 1 and Half 2. Aside from minor differences in the extreme values, there is a high degree of consistency across the figures.

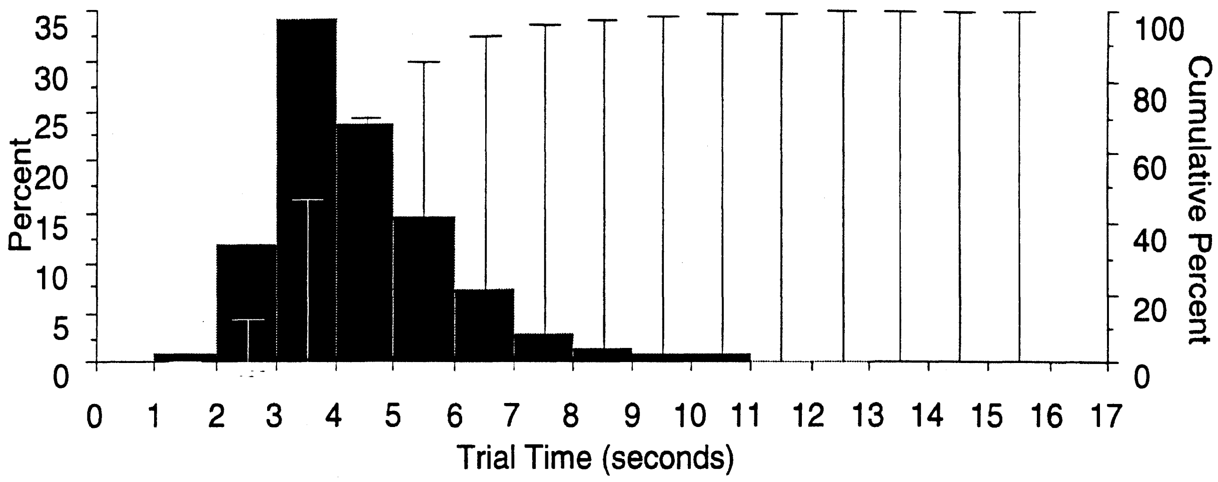


Figure 12. Histogram of time for all the data.

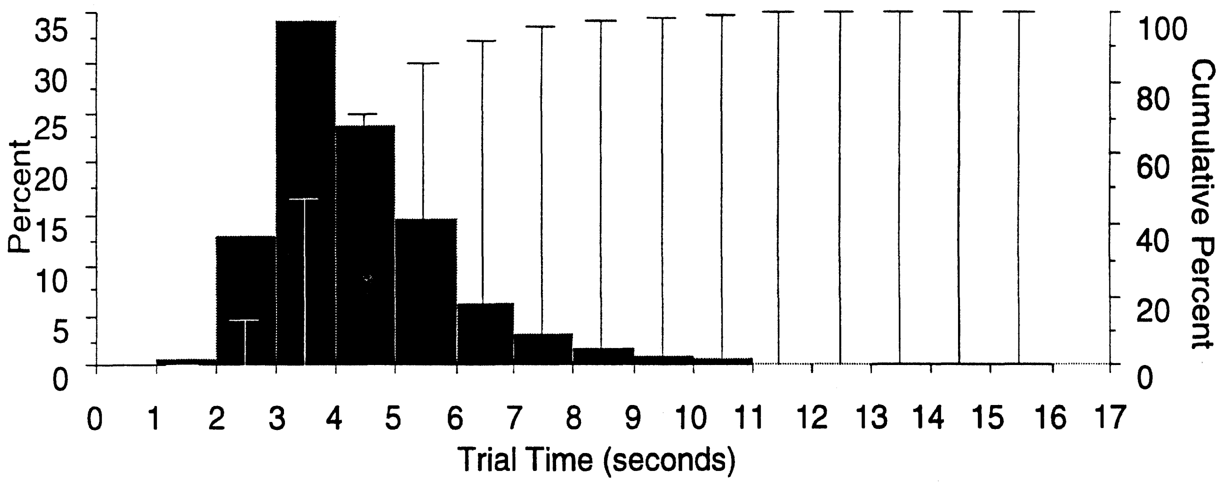


Figure 13. Histogram of time for Half 1.

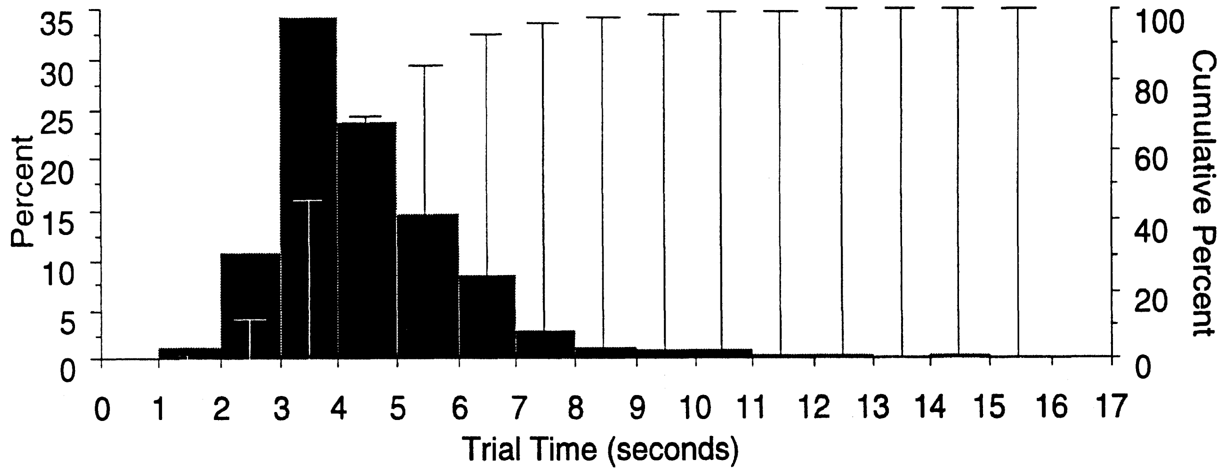


Figure 14. Histogram of time for Half 2.

The second comparison involved performing two sets of repeated-measures ANOVAs. Each set consisted of one ANOVA for all the data and one for each of the two halves. The model for the first set contained menu structure, control, configuration, and all interactions between them, while that of the second contained simply the block number (in order to evaluate the learning effect). Age and sex were not included in either of the models because all four participants were from the same age group and only one of them was female. Finally, time was the sole dependent measure. (Errors were not analyzed because there were only 27 errors across the four subjects, an average of just over half an error per block).

The conclusions that can be reached for these ANOVAs are quite similar (see Table 9). The effect of configuration and menu structure were insignificant no matter which set of data was examined. The effect of control was significant at the 0.05 level for all the data and for Half 1 and, at the 0.1 level for Half 2. The block effect was insignificant at the 0.05 level for all three ANOVAs and at the 0.1 level for two of the three. Finally, the p-values were relatively consistent across the three sets of data except for the effect of menu structure for Half 2 which was considerably higher.

Table 9. ANOVA results for each subset of the data.

		P-Value		
Source	df	All	Half 1	Half 2
Configuration	2	.73	.80	.69
Control	1	.044*	.025*	.073
Menu Structure	1	.59	.50	.86
Block	11	.12	.13	.092

Note: P-values were adjusted via the Huynh-Feldt method.
* p < .05

A final repeated-measures ANOVA was performed whose model included menu structure, control, configuration, data included, and all interactions between them. The key result is that the data-included factor (which measured the difference between using all the data, Half 1, and Half 2) was insignificant (p=0.26). Thus, all of the data and either half of the data led to the same conclusions with regard to the significance of experimental variables..

Implications

In conclusion, the cursor control took less time than the number pad and examining only the data corresponding to 16-item prompt sets produced essentially the same results as examining all the data. Hence, the cursor control and 16-item prompt sets were used in the main experiment.

MAIN RESULTS

Overview

The description of the results of the main experiment is divided into six sections. The first deals with quantitative results concerning the menu task and is the most comprehensive. Discussion includes typical selection times and error rates and the effect of each experimental factor on time, accuracy, and various fine-grained keystroke measures. The second section describes similar issues for some of the driving performance measures logged by the simulator, namely lane excursions per block, standard deviation lane position, and standard deviation speed. The third section details the creation and verification of a GOMS (Goals, Operators, Methods, and Selection rules) model (Card, Moran, and Newell, 1983) for the in-vehicle menu system tested in this experiment.

The final three sections are more qualitative in nature. The fourth discusses the results of the acceptability survey, which indicates how the drivers perceived the in-vehicle menu system. The fifth and sixth sections concern experimenter observations and include the results of the informal eye-fixation analysis and discussion of how subjects reacted when introduced to the interface.

Menu task results

The MenuPlayer output log files contained the trial time and accuracy, the number of extra (unnecessary) keystrokes involved in each trial, and time stamps for all keystrokes. These files were concatenated and spreadsheets were created summarizing the data for each block of trials for each subject.

From these files, five dependent measures were computed: time, accuracy, extra keystrokes, up keystrokes, and back keystrokes. For some of these measures, analyses were performed with specific subsets of trials—*correct* trials and *ideal* trials —because they provided more meaningful information. Correct trials are where the user's final selection was correct, while ideal trials are correct trials where the back button was not used. Ideal trials are assumed to be performed more or less expertly, as no trial and error was involved in finding the correct menu items and the task was completed successfully. Table 10 contains a complete description of each dependent measure.

Repeated-measures ANOVA models were used to examine each of these dependent measures since both between-subject and within-subject factors were present. One model contained age, sex, menu structure, control, configuration, and all interactions between them (except those involving both age by sex). A second model tested the learning effect and contained age, sex, block, block x age, and block x sex. Table 11 shows the factors explored and their significance levels for each dependent measure. A detailed analysis of each factor in these models follows. In brief, the general pattern was that age and block differences were consistently significant, the effect of control and block were occasionally significant, and the remaining factors were rarely significant.

Table 10. Summary of the five dependent measures.

Measure	Description	Trials Analyzed	Rationale
Time	Time (in seconds) to complete a single trial	All	Provides the true time needed to complete a task, indicating how well novices perform
		Ideal*	Estimates task times for experienced users
Accuracy	Correctness of a trial (either 1 or 0)	All	Yields error data for each trial
Extra keystrokes	Number of unnecessary button presses or knob turns per trial	Correct	Indicates how efficiently and directly each trial was completed
Up keystrokes	Number of up button presses per trial where the selection bar was not already at the top	All	Indicates the number of pure scroll errors per trial (i.e. how many times and to what extent the selection bar scrolled beyond the desired item)
Back keystrokes	Number of back button presses per trial where the main menu was not already showing	All	Reveals the number of times backtracking occurred per trial, indicating either uncertainty about the location of menu items or accidental keypresses

* Ideal trials are correct trials where the back button was not used.

Table 11. ANOVA results for the menu task.

	P-Value					
	Primary Measures			Keystroke Measures		
Factor/ Interaction	Trial Time	Time (Ideal)	Error Rate	Extra (Correct)	Up	Back
Age	< .001*	< .001*	< .001*	.78	.039*	<.001*
Sex	.018*	.055	.25	.58	.21	.32
Menu structure	.59	.91	.77	.38	.027*	<.001*
Control	.049*	.32	.82	.013*	.0013*	.0022*
Configuration	.41	.14	.68	.66	.61	.80
Block	< .001*	< .001*	.010*	.078	.42	.031*
Control x Age	.018*	.063	.81	.99	.85	.053
Control x Sex	.46	.90	.018*	.64	.65	.38
Control x M.S.	.09	.67	.87	.16	.017*	.072
Block x Age	.0054*	.11	.089	.50	.46	.52

* p < .05

Typical selection times and error rates

Across all subjects, the mean time to complete a trial was 9.14 seconds, with times ranging from 1.27 to 87.28 seconds. For ideal trials, the mean time fell to 7.63 seconds and the maximum to 49.08 seconds. The histograms below (see Figures 15 and 16) reveal that the difference between the two means was due primarily to a reduction in the number of long times and outliers. A likely explanation is that, for ideal trials, the driver knew (or correctly guessed) which menu items to choose and did not spend time backtracking.

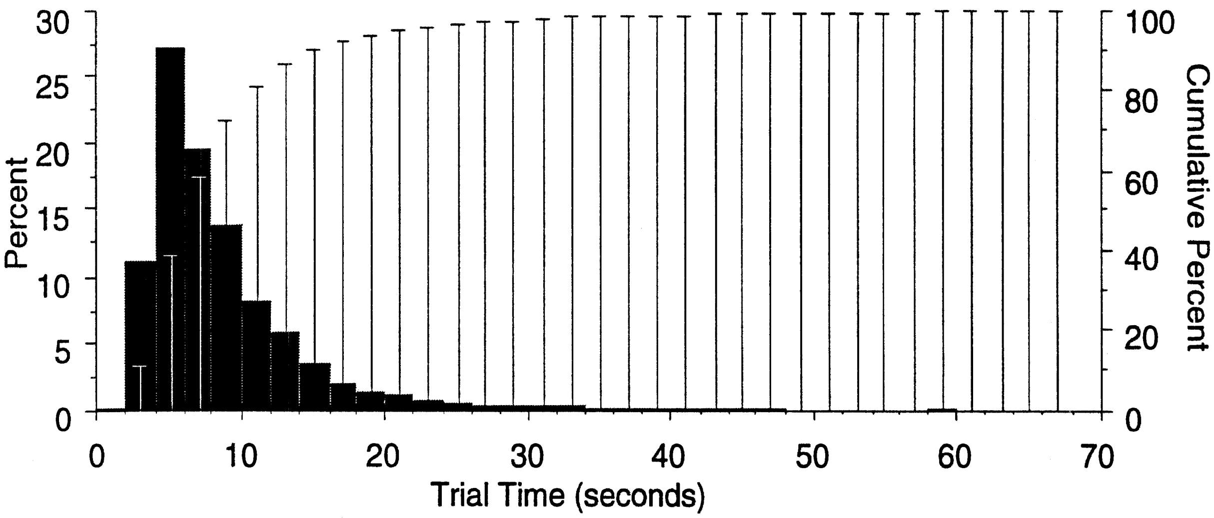


Figure 15. Histogram of time for all trials.

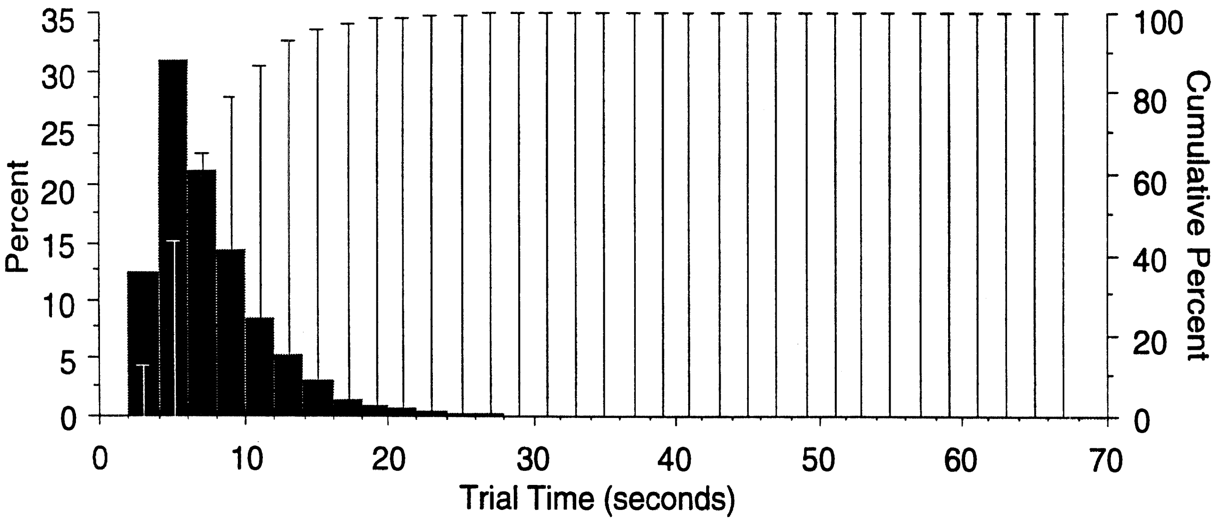


Figure 16. Histogram of time for ideal trials.

The mean error rate across all participants was 7.8 percent. Errors were generally due to the driver either (1) selecting an incorrect item thinking it was correct, (2) selecting an incorrect item being unable to find the correct one, or (3) selecting an item by accident (slipping).

The mean number of extra keystrokes for correct trials was 0.96. The maximum was 35. Figure 17 reveals that the mode was zero (71 percent of all trials) and values above 10 were very rare. Interestingly, there were more trials involving even numbers of extra keystrokes than odd because each movement of the selection bar past the desired item requires a corrective movement in the opposite direction. Overall, the extra keystroke data suggests that subjects were quite efficient, with the vast majority of correct trials involving two or less unnecessary button presses or knob movements.

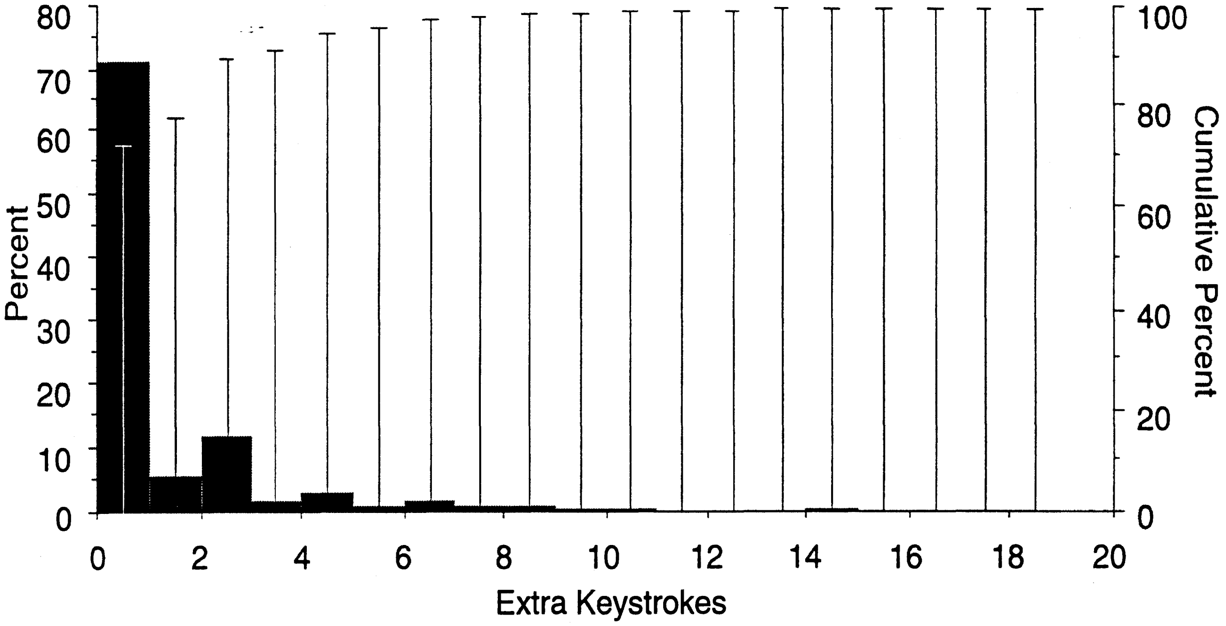


Figure 17. Histogram of extra keystrokes for correct trials.

The number of up keystrokes per trial indicates how many extra keystrokes were purely a result of scrolling past the desired menu item and scrolling back up. The mean value was 0.24, suggesting that a little more than one in five trials involved a scrolling error. The maximum was 15, perhaps indicative of confusion or carelessness.

Back keystrokes occurred when an item was accidentally selected or when there was uncertainty about which menu selections were required to perform the desired task. The mean number of back keystrokes per trial was 0.11 and the maximum was five. Hence, approximately one in ten trials involved backtracking, although the exact cause cannot be determined. Thus, within-menu (scrolling) errors were more than twice as prevalent as between-menu (backtracking) errors.

Effects of age and sex

As noted previously, age and sex effects were observed for several of the dependent measures. The effect of age was significant for all dependent measures except extra keystrokes. As Table 12 indicates, when all trials were examined, older subjects took almost double the time of young subjects to complete a trial. When only ideal trials were analyzed, the means dropped to 9.81 and 5.77, respectively, but the difference was still highly significant (p < .001). These differences in time can be attributed to a variety of cognitive and physiological factors, such as memory capacity and hand-eye coordination. Knowledge and experience were likely important factors as well. For example, computer usage varied greatly with young subjects averaging 31.6 hours per week and older participants averaging just 4.8.

Table 12. Results concerning the effects of age and sex.

		Time (seconds)		Error Rate (%)	
Factor	Level	Mean	P-Value	Mean	P-Value
Age	Young	6.16	< .001*	2.3	< .001*
	Older	12.12		13.3	
Sex	Men	7.87	.018*	6.2	.25
	Women	10.41		9.3	

* p < .05

Young subjects were significantly more accurate than older ones (by a factor of more than five). From a customer perspective, the 13 percent error rates found for older subjects are likely to be unacceptable. Also, older subjects made more accidental selections and had more difficulty learning and remembering the location of certain menu items than younger subjects.

Although there was no significant difference among age groups for extra keystrokes (p=0.78), there were significant differences for the more specific up and back keystroke measures (p=0.039 and p<0.001, respectively). Older participants made fewer scrolling errors but engaged in more backtracking than young subjects (see Figure 18). The higher number of scrolling errors for young subjects suggests that they may have worked faster at the expense of accuracy, while older participants may have taken more care in operating the controls. The increased incidence of backtracking among older participants may suggest that they had more difficulty remembering where certain items were in the menu structure.

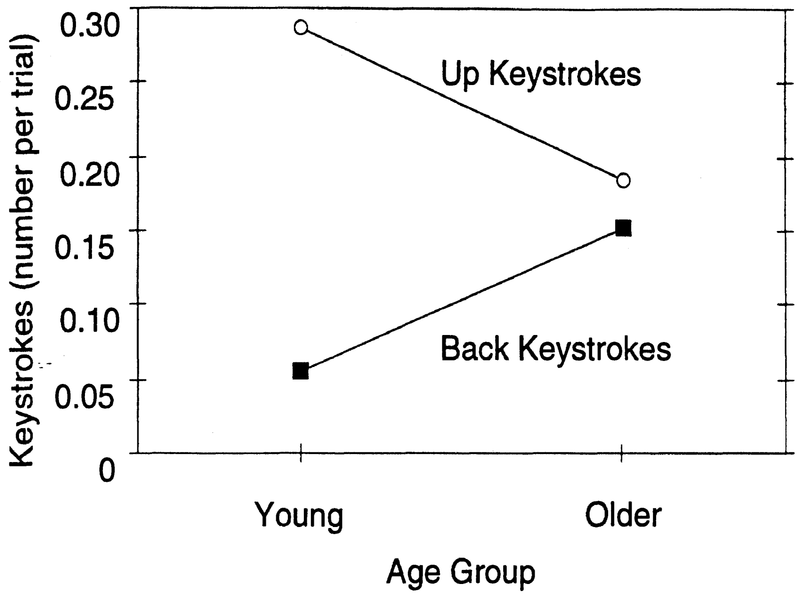


Figure 18. The effect of age on up and back keystrokes.

The sex effect was significant for time only when all trials were analyzed. Women took 32.2 percent longer on average to complete a trial than men. There was no significant difference in times when only ideal trials were considered (p=0.059), however. There was also no significant difference between men and women in accuracy, extra keystrokes (p=0.85), up keystrokes (p=0.21) or back keystrokes (p=0.32). This may imply that women spent more time thinking about which items belonged to which menus, possibly due to less familiarity with the automotive terms used.

Effects of menu structure, control, and configuration

As Table 13 indicates, menu structure did not have a significant effect on time or error rate when all trials were examined. The similarity in trial times may be the result of a balancing effect in which the four extra items per menu for the broad structure had the same effect on trial times as the extra menu for the deep structure. The effect of menu structure also had no significant impact on time when only ideal trials (p=0.91) or extra keystrokes (p=0.38) were examined.

Table 13. Results concerning the effects of menu structure, control, and configuration.

		Time (seconds)		Error Rate (%)	
Factor	Level	Mean	P-Value	Mean	P-Value
Menu structure	Broad	9.05	.59	7.9	.77
	Deep	9.23		7.7	
Control	Cursor	9.45	.049*	7.9	.82
	Knob	8.87		7.7	
Configuration	Both-High	9.06	.41	8.2	.68
	Both-low	9.49		7.2	
	Separated	8.88		7.9	

Note: P-values were adjusted via the Huynh-Feldt method.
* p < .05

The menu structure did, however, significantly effect the number of up and back keystrokes per trial (p=0.027 and p<0.001, respectively). The mean number of up keystrokes was greater for the broad structure than for the deep structure, while the mean number of back keystrokes was greater for the deep structure than the broad structure (see Figure 19). The broad menu structure yielded more scrolling errors most likely because the menus contained eight items each, requiring more scrolling, hence increasing the chance of scrolling errors. On the other hand, the broad menu structure resulted in less backtracking probably because it required fewer menu selections and displayed more options at a time, reducing the need for the user to guess.

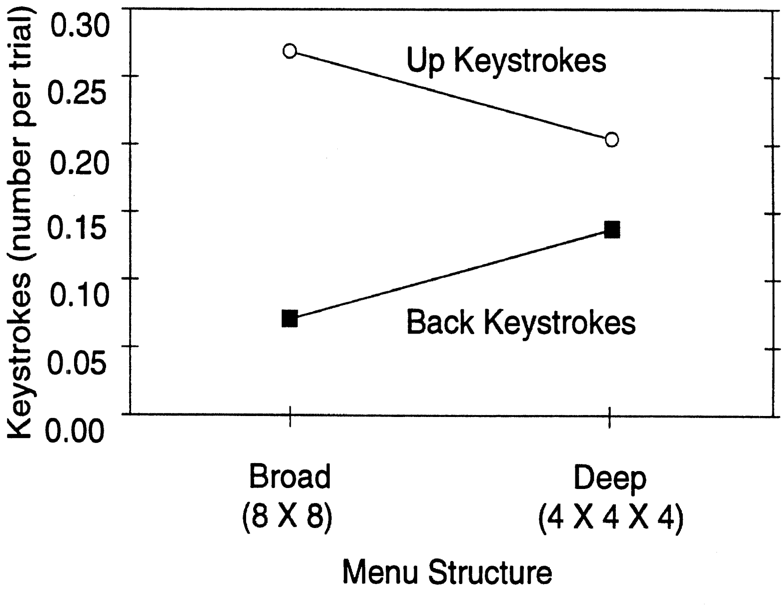


Figure 19. The effect of menu structure on up and back keystrokes.

The type of control had a significant effect on task time when all trials were examined. Specifically, trials involving the knob control were completed over half a second faster than those with the cursor control. The 95 percent confidence interval for the difference between the two means is 0.27 to 0.82. When only ideal trials were examined, however, the mean difference shrank to 0.2 seconds and the effect was no longer statistically significant (p=0.32). This may be because non ideal trials involving the back key involved more keystrokes than ideal trials, allowing the differences between the controls to become more pronounced.

Examining the effect of control and age on time (see Figure 20) revealed that the knob control was faster only among the older participants, as the younger subjects were actually slightly quicker with the cursor control. This may again be a result of computer experience, as the cursor control was typical of what would be found on a computer keyboard. This interaction was not significant for ideal trials (p=0.063).

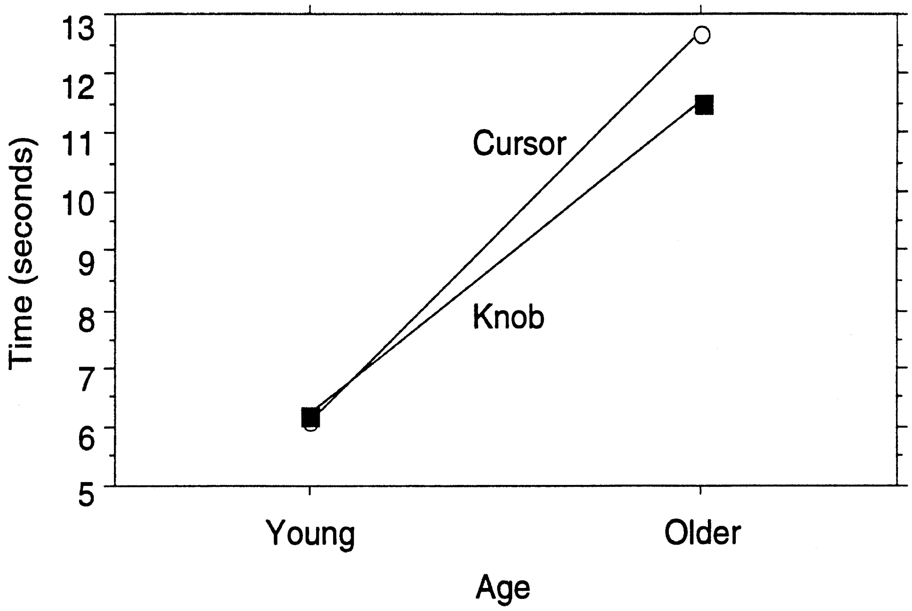


Figure 20. The effect of control and age on time for all trials.

Although there was no main effect of control on accuracy, there was a significant interaction between control and age (p=0.018). Men committed less errors using the cursor control while women were more accurate with the knob control (see Figure 21). In either case, however, the mean difference in error rates between the controls did not exceed 2.1 percent.

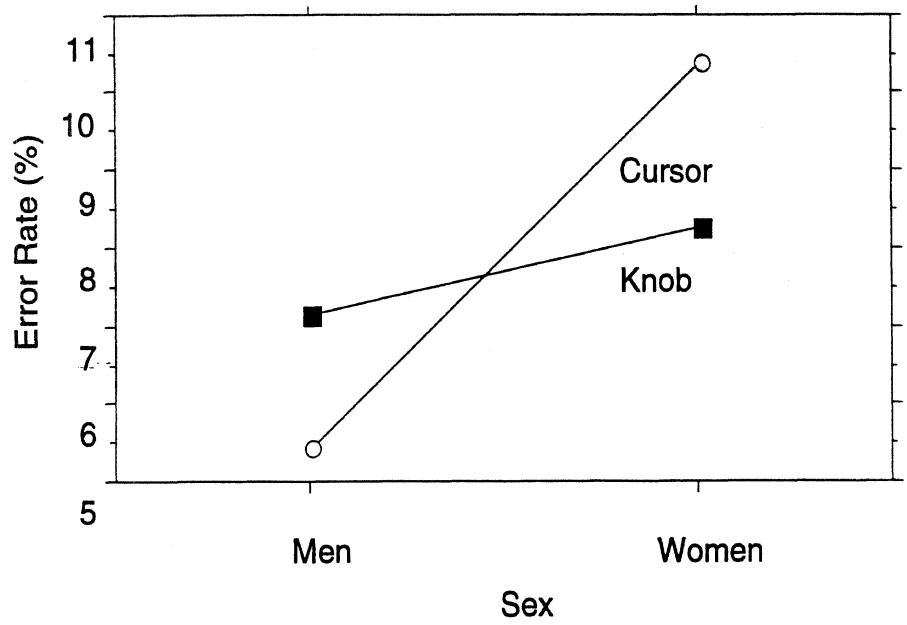


Figure 21. The effect of control and sex on error rate.

The difference between the cursor and knob controls was significant for each of the three keystroke measures: extra keystrokes (p=0.013), up keystrokes (p=0.0013), and back keystrokes (p=0.0022). As Figure 22 shows, the knob yielded more extra keystrokes and up keystrokes than the cursor, while the cursor yielded more back keystrokes. The differences between the means, however, were no greater than 0.1 seconds.

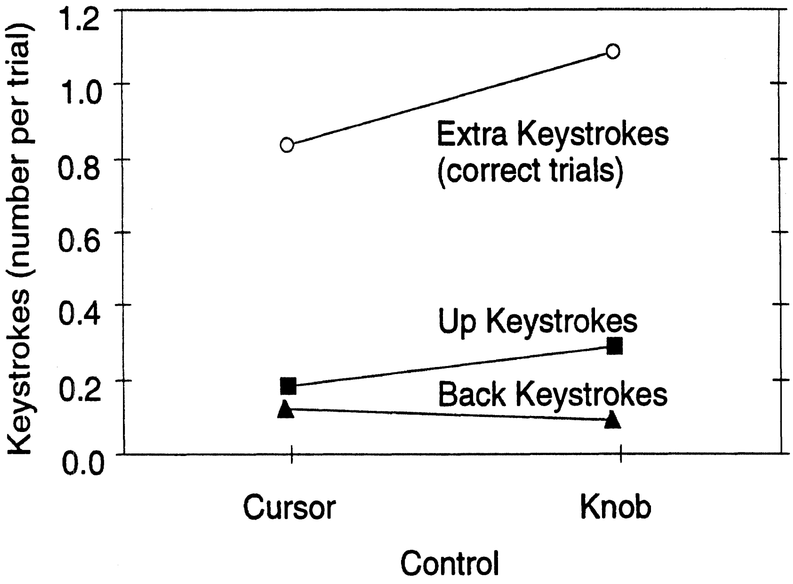


Figure 22. The effect of control on extra keystrokes for correct trials and up and back keystrokes for all trials.

The control/display configuration did not have a statistically significant effect on time when all trials were examined. The means suggest that having the control high on the dashboard and the control comfortably in a lower position may allow for the fastest times, although these results could easily be due to chance. The configuration also had no significant effect on time when only ideal trials (p=0.14), accuracy, or any of the keystroke measures (p=.66, .61, and .80 for extra keystrokes, up keystrokes, and back keystrokes, respectively) were examined.

Learning effect

The learning effect was evaluated by comparing the 12 blocks of the experiment for each of the four dependent measures. There was a significant block effect for time (see Table 14) and a significant interaction between block and age (p=0.0054, adjusted via the Huynh-Feldt method) when all trials were included. As Figure 23 indicates, older subjects showed considerably more improvement from the first to the last trial than young ones (5.45 seconds versus 2.01). This may be because the older subjects had more room for improvement than the young ones, perhaps due to the lack of computer experience.

Table 14. Results concerning the learning effect.

		Time (seconds)		Error Rate (%)	
Factor	Level	Mean	P-Value	Mean	P-Value
Block	First	11.24	< .001*	11.2	.010*
	Last	7.50		5.7	

Note: P-values were adjusted via the Huynh-Feldt method.
* p < .05

A decreasing trend was also present when only ideal trials were considered (p<0.001), indicating that subjects improved their time even on trials where they chose only correct menu items (see Figure 24). The slight variation in the data is probably random noise.

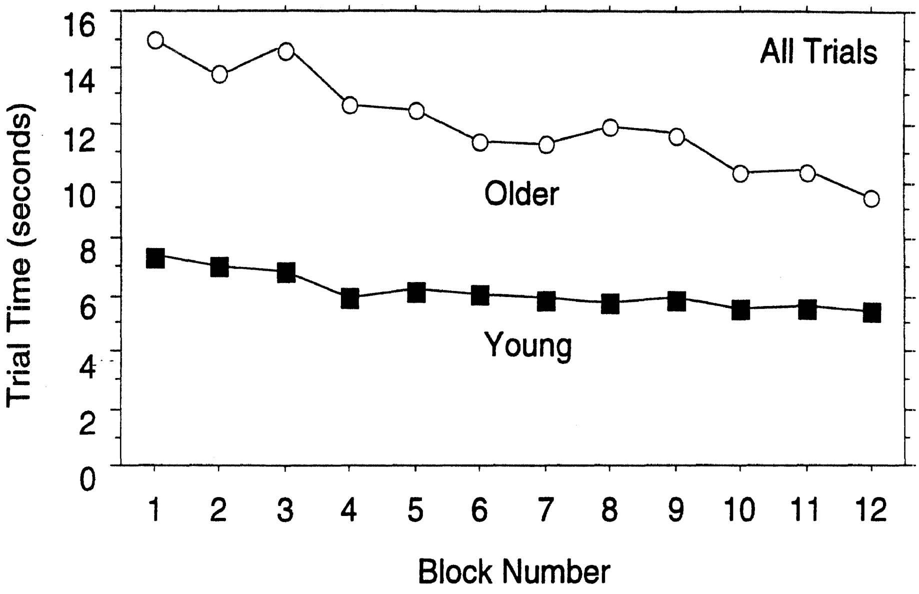


Figure 23. The effect of block and age on time for all trials.

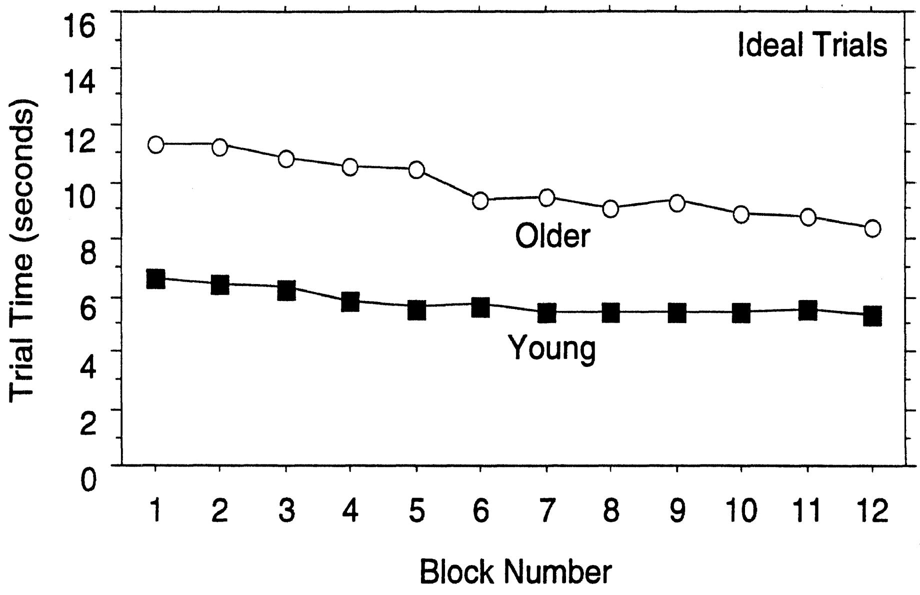


Figure 24. The effect of block on time for ideal trials.

The learning effect was also present in the accuracy data (see Figure 25). The large amount of variability could be a result of the relatively small number of errors overall.

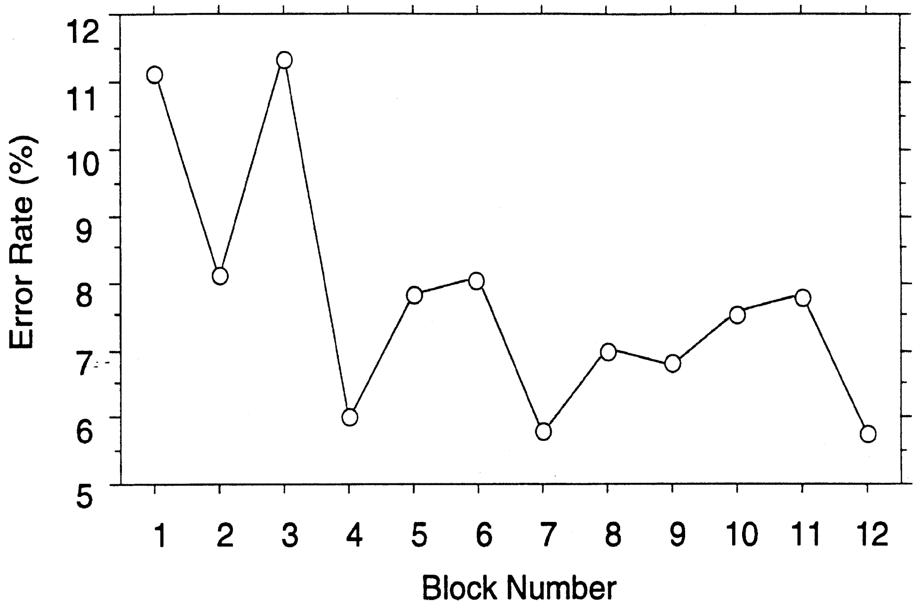


Figure 25. The effect of block on error rate.

Among extra keystrokes, up keystrokes, and back keystrokes, the block effect was only significant for back keystrokes (p=0.031, adjusted via the Huynh-Feldt method). Except for the eighth block (an apparent outlier), there was a general decreasing trend (see Figure 26). This indicates that participants became increasingly familiar with the database and had less of a need to engage in backtracking as the experiment progressed. The fact that up keystrokes did not differ significantly among the blocks may suggest that subjects put little effort into minimizing scrolling errors.

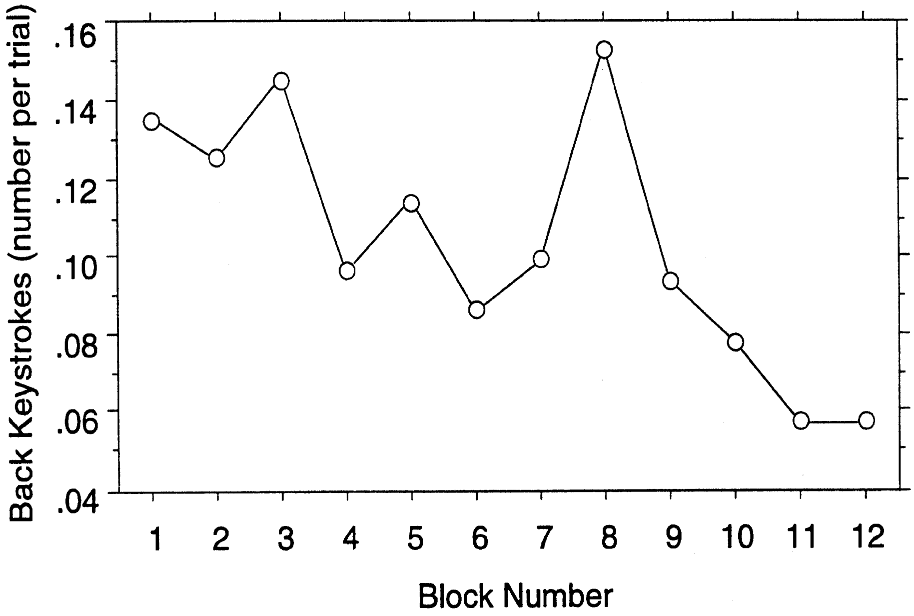


Figure 26. The effect of block on back keystrokes.

Driving performance

To help evaluate the effect of the in-vehicle menu interface on driving performance and safety, the data logged by the driving simulator was analyzed. The first 30 seconds of data were eliminated from each file because speeds were still stabilizing during that period. Also, one subject was omitted from the analysis because a file was lost.

The dependent measures included (1) lane excursions per block (i.e., instances where the car crossed over a lane marker), (2) standard deviation lane position (in feet), and (3) standard deviation speed (in miles per hour). Two repeated-measures ANOVAs were performed for each dependent variable. The first involved age, sex, menu structure, control, configuration, and all interactions between them (except those involving both age by sex). The second involved age, sex, block, block x age, and block x sex. Table 15 contains a summary of the factors and significant interactions and corresponding p-values for each of the three dependent measures.

Table 15. ANOVA results for the driving data.

	P-Value		
Factor/Interaction	Lane Excursions	SD Lane Position	SD Speed
Age	.028*	.44	.30
Sex	.93	.22	.56
Menu Structure	.055	.23	.14
Control	.94	.39	.03*
Configuration	.95	.36	.98
Block	.0017*	.34	.07
Control x Sex	.26	.40	.0042*

* p < .05

Typical values

The mean number of lane excursions per block across all subjects was 5.74. This indicates that the task contributed to approximately one lane excursion every two minutes since each block of 16 trials generally took under 10 minutes. Furthermore, as Figure 27 shows, over 80 percent of blocks involved at least one lane excursion.

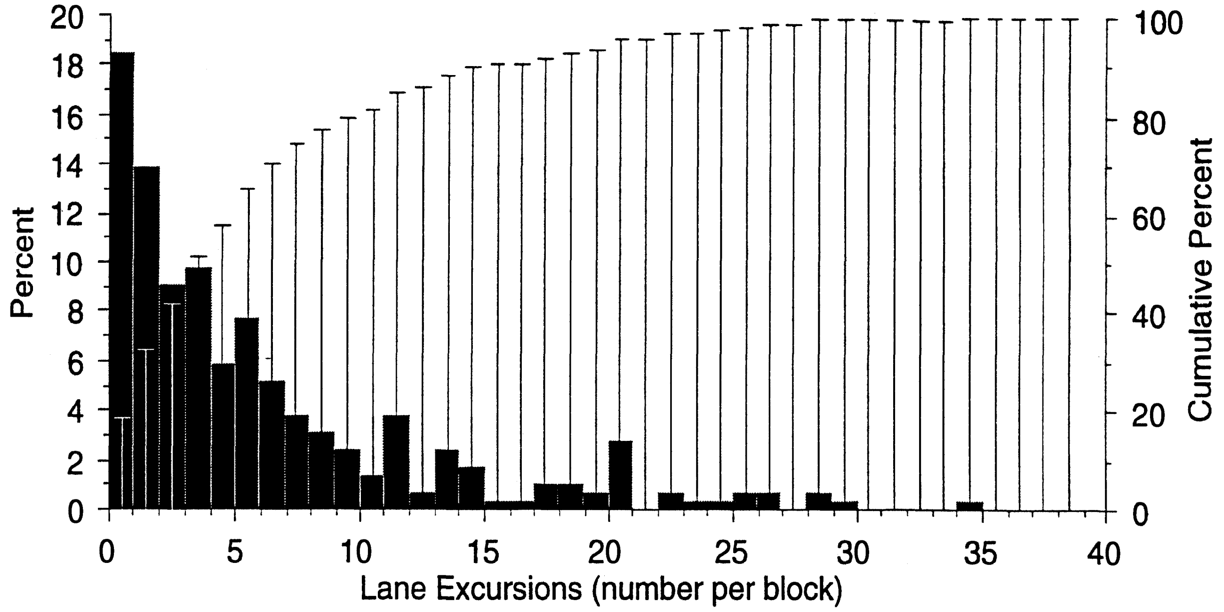


Figure 27. Histogram of lane excursions per block.

Across all subjects, the mean standard deviation lane position and mean standard deviation speed were 1.36 feet and 5.28 miles per hour, respectively. The distributions for both were fairly normal (see Figures 28 and 29). Also, both means were higher than those typically observed for driving alone. For example, Green (1995) noted that, for driving simulators, 0.39 to 0.59 feet was typical for standard deviation lane position and 1.0 to 2.0 miles per hour was typical for standard deviation speed. These values were derived from a number of studies where different types of roads were used.

In a more recent study involving an instrumented test vehicle driven over a variety of roads, Katz, Fleming, Green, Hunter, and Damouth (1997) report a range of 0.7 to 1.2 feet for standard deviation lane position and a span of three to 12 miles per hour for standard deviation speed. The on-road speed variance was much greater than that found in the laboratory because traffic lights and other vehicles prevented drivers from maintaining constant speeds. On the other hand, on-road lateral variance was greater in the simulator because there were more curves. Although identifying a relevant comparison baseline is difficult, the simulator data suggest the menu interface used in this experiment may have been responsible for a degradation in steering ability and speed maintenance.

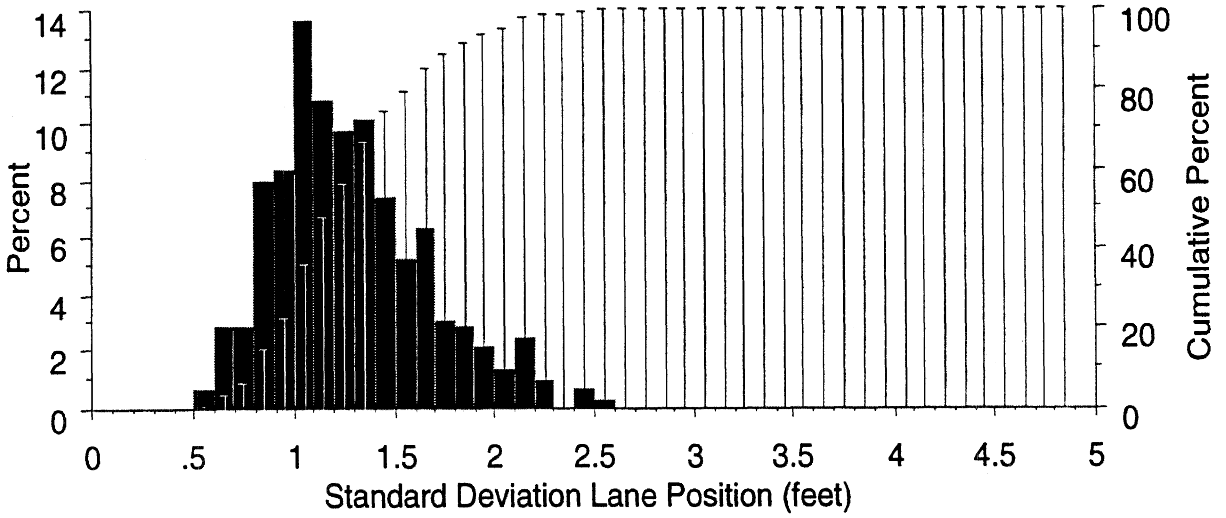


Figure 28. Histogram of standard deviation of lane position.

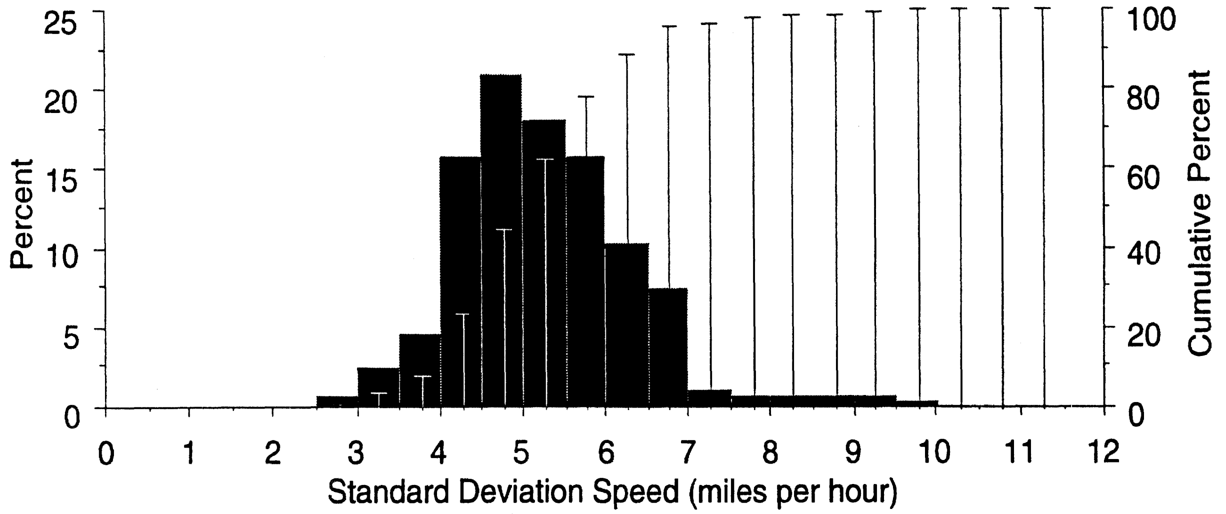


Figure 29. Histogram of standard deviation of speed.

Means and ANOVA results

Table 16 contains a summary of the effect of each within- and between-subject factor on the three driving measures. Most notably, older subjects exceeded lane boundaries more than 2.5 times as often as young ones did. Also, the number of lane excursions fell by 27 percent from the first block of trials to the last (see Figure 30). Standard deviation speed was significantly higher for the cursor control than the knob control but the difference was only 0.32 miles per hour. There were also two significant interactions concerning standard deviation speed, control x sex (p=0.0042) and configuration x menu structure (p=0.021), but neither appeared to have any practical significance. Finally, there were no significant main effects or interactions for standard deviation lane position.

Table 16. Results concerning driving performance.

		Lane Excursions (no. per block)		SD Lane Position (feet)		SD Speed (miles per hour)	
Factor	Level	Mean	P-Value	Mean	P-Value	Mean	P-Value
Age	Young	2.87	.028*	1.27	.44	7.92	.30
	Older	8.03		1.43		7.46	
Sex	Men	5.35	.93	1.47	.22	7.82	.56
	Women	5.32		1.21		7.57	
Menu structure	Broad	5.67	.055	1.45	.23	7.83	.14
	Deep	5.00		1.24		7.57	
Control	Cursor	5.32	.94	1.42	.39	7.86	.027*
	Knob	5.36		1.27		7.54	
Configuration	Both-High	5.34	.95	1.52	.36	7.68	.98
	Both-low	5.21		1.26		7.70	
	Separated	5.47		1.26		7.71	
Block	First	6.22	.0017*	1.24	.34	7.13	.071
	Last	4.52		1.26		8.42	

Note: P-values for within-subject factors were adjusted via the Huynh-Feldt method.
* p < .05

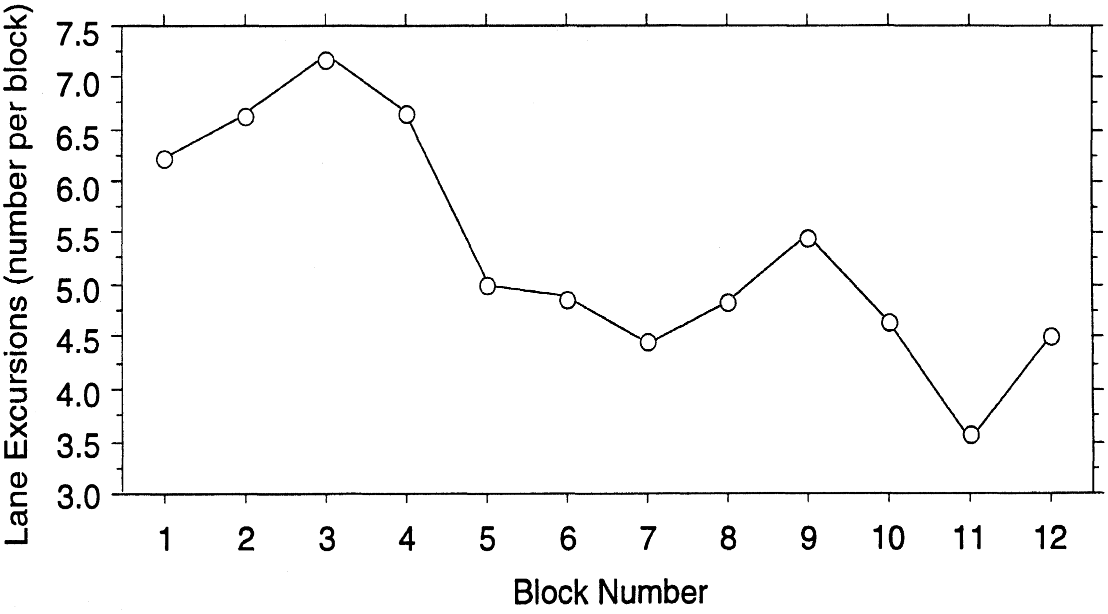


Figure 30. The effect of block on lane excursions.

GOMS analysis

In addition to these primary analyses, a GOMS analysis was performed to determine whether task times for menu interfaces can be predicted using individual keystroke times and simple formulas. GOMS modeling entails (1) dividing a task into its component steps, (2) obtaining a time for each step, and (3) calculating the total task time using a function of the step times (Card, Moran, and Newell, 1983). The value of such a model is that it can be used to predict task times for interfaces before they are built, saving development time and the cost of building unnecessary prototypes. This section describes the generation of a GOMS model for the interface used in this experiment and discusses how accurately the model predicted the actual trial times.

The first phase of the modeling process was to create a flowchart to determine the key steps involved in completing a trial (see Figure 31). As is typical with GOMS modeling, errors were not taken into account in order to keep model complexity to a minimum. Also, the first step, listening to the request, was not included in the model because its time was merely a function of the digitized sound clips.

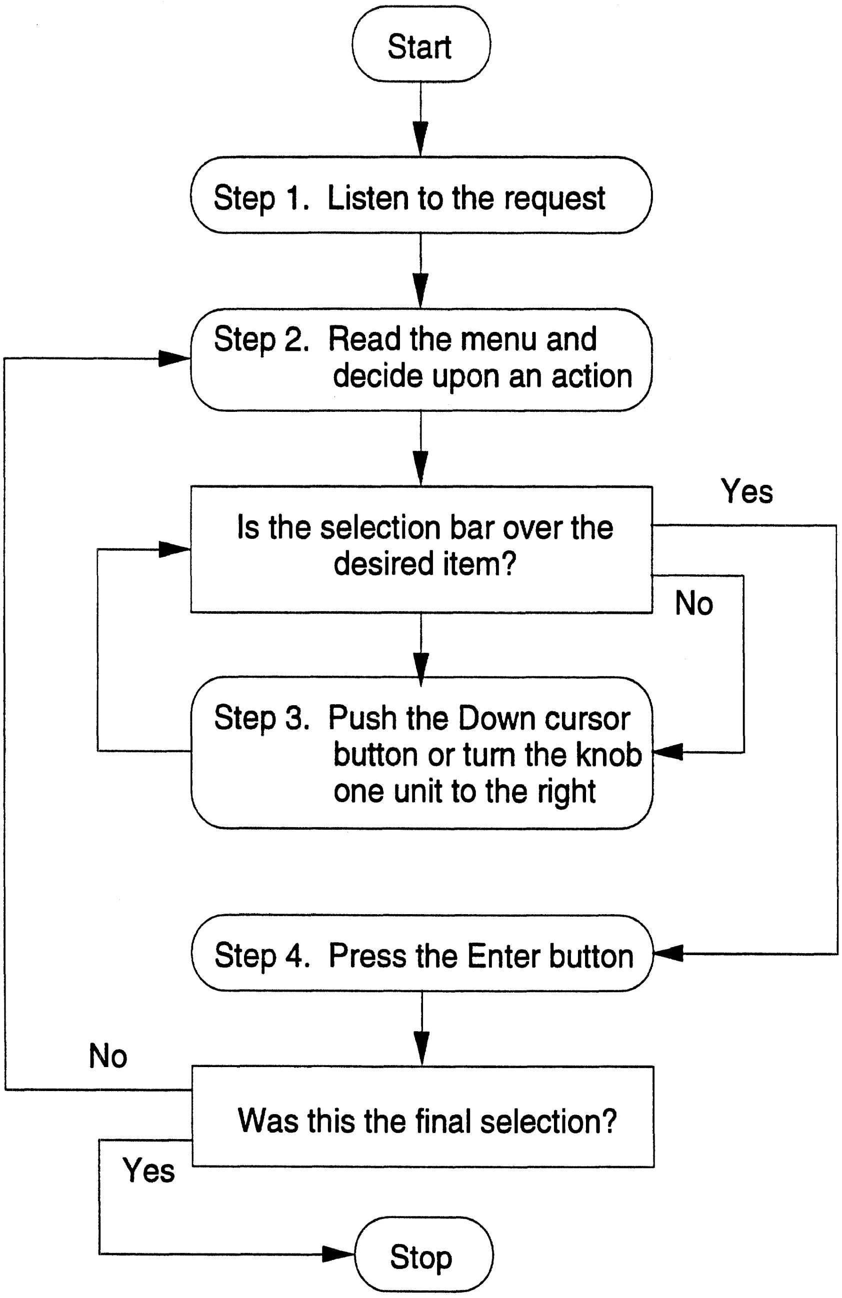


Figure 31. Flowchart showing the steps needed to complete each trial.

The time for each step was determined using the experimental data. This involved calculating the mean times for down and enter keystrokes (steps 3 and 4) and deriving the time for reading the menu and deciding upon an action (step 2). Because the possibility of errors was not taken into account, only data from correct trials involving no extra keystrokes were used in these calculations. Also, to minimize variability caused by confusion and large between-subject differences, only the data from the young subjects was used (see Figure 32).

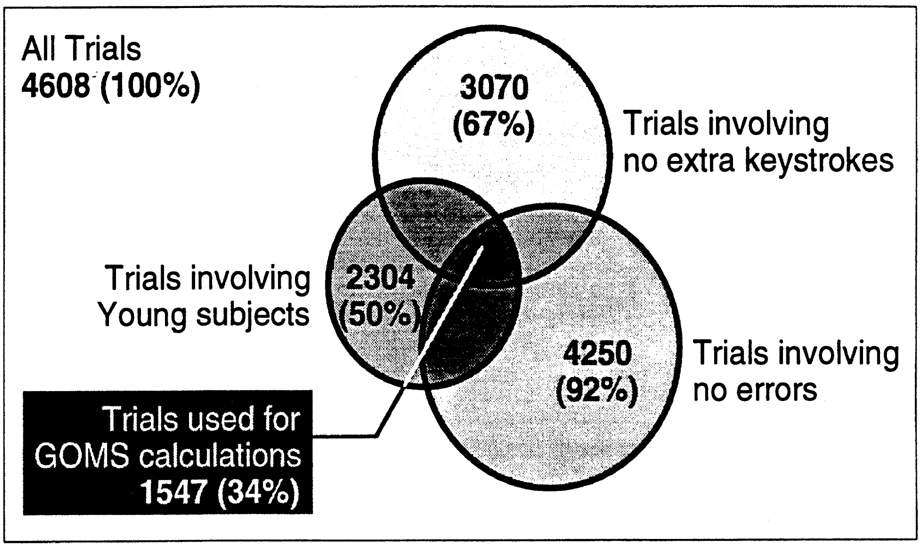


Figure 32. Venn diagram showing the number and type of trials used in calculating the GOMS parameters.

Figure 33 shows the time to execute a down keystroke as a function of the current position of the selection bar for each menu structure. The first time in each graph consists not only of the time to press the down button, but also the time to read the menu and decide upon an action. Therefore, the average times to execute a down keystroke (0.29 and 0.28 seconds for the broad and deep menu structures, respectively) were based only on instances where the selection bar was beyond the first item. Finally, it appears that there was no meaningful difference between the two controls and that times did not change substantially as subjects scrolled from the first item to the last.

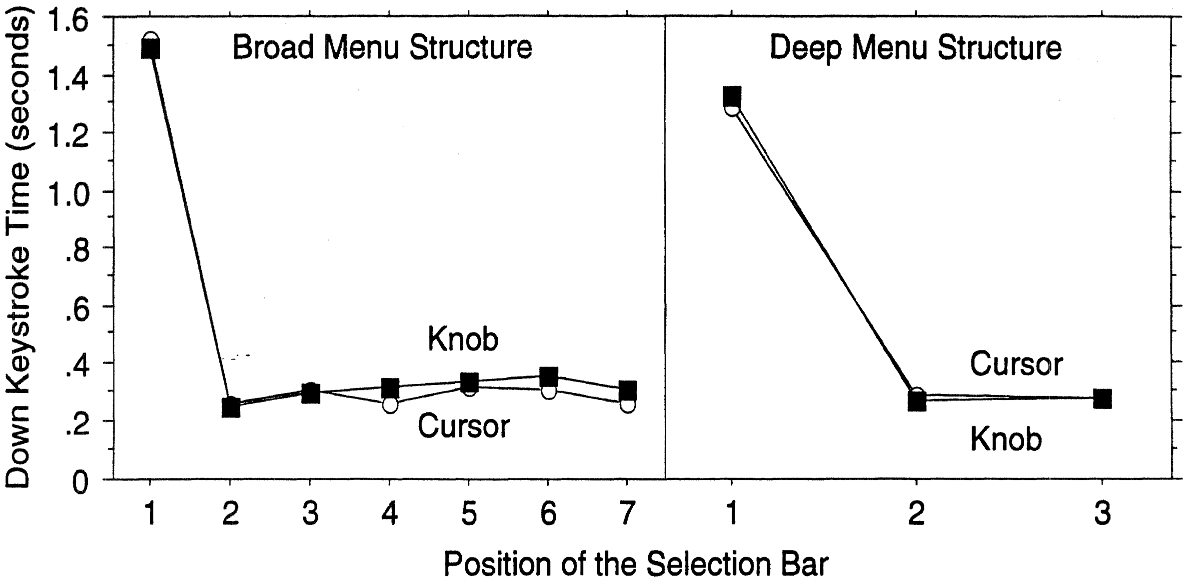


Figure 33. The effect of the position of the selection bar on the down keystroke time.

The plots for the enter keystrokes (see Figure 34) are similar to those for the down keystrokes. As before, the first times included mental operations as well as the time to execute the keystroke, while the remaining times were much closer to pure keystroke times. Unlike the case of the down keystrokes, however, it appears that there were nontrivial differences between the two controls as pressing the enter button took longer with the knob control than with the cursor. With the knob control, users had to remove one of their fingers from the knob to find the enter button, while with the cursor control, their finger was generally already in the proper position. The average times for each condition are presented in Table 17.

Table 17. The average enter keystroke time by menu structure and control.

	Extra Keystroke Time (seconds)		
Menu Structure	Cursor Control	Knob Control	Difference
Broad	.45	.57	.12
Deep	.38	.47	.09
Difference	.07	.10	

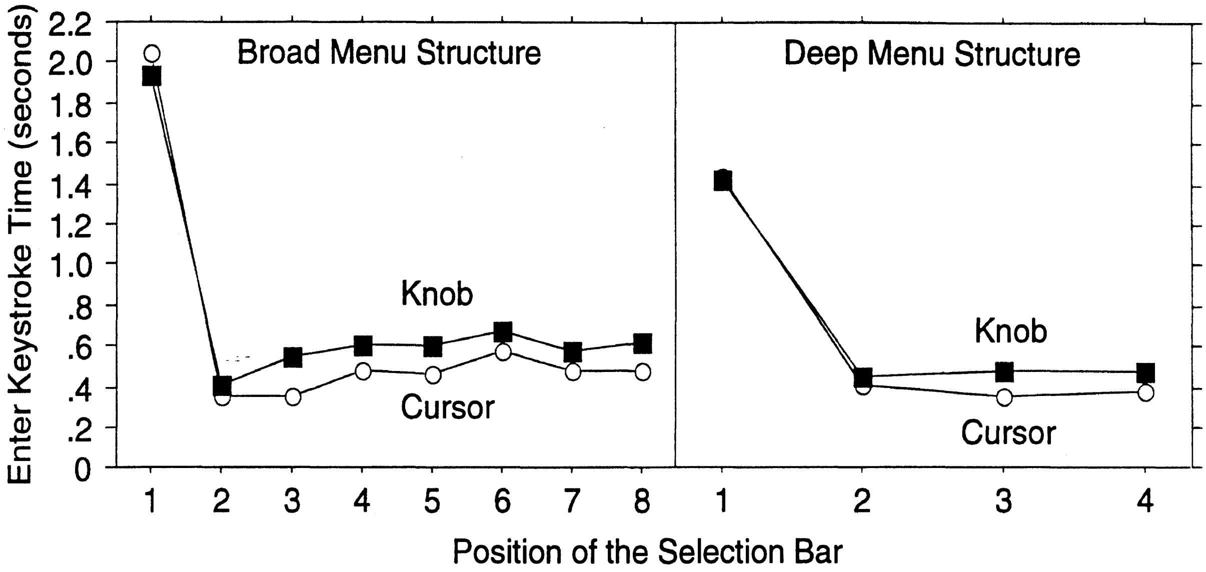


Figure 34. The effect of the position of the selection bar on the enter keystroke time.

The time for reading the menu and deciding upon an action (referred to as the reading/decision time) was derived by subtracting the average down and enter keystroke times from the times for the first keystroke of each trial (which contained both the reading/decision time and the time to press the down or enter button). A graph of the reading/decision time as a function of the target menu item was produced for each menu structure (see Figure 35). These provided information about how the position of the desired item in the list affected the time spent reading the menu.

The graph for the broad structure reveals a trend that is far from linear. For both controls, the reading/decision time was the highest for menus where the first item was the target. This is probably because most menus involved scrolling before selecting the item, so when scrolling was not required, extra time was likely taken to resist the impulse to do so. For the cursor control, the reading/decision time was shortest for menus where the last item was selected, perhaps indicating that subjects began scrolling before reading all the choices (hence shortening the initial reading/decision time) or that they glanced to the bottom of the list before reading the earlier choices. Finally, for both controls, there was an inexplicable alternating trend, although it was less pronounced for the cursor control.

The graph for the deep menu structure is somewhat different. First, the reading/decision times for menus where the first item was the target were no longer the longest. This is possibly because the first item was selected much more frequently with the deep structure than with the broad, allowing the subject to get used to the slight difference in the procedure. A second difference is that the times for menus where the last item was selected were among the longest for the deep structure, while they were among the shortest for the broad. This could mean that subjects used a different reading strategy for the two structures. For example, they might have read the items in order for the deep structure but skipped around more for the broad structure. Finally, the overall reading/decision times for the deep menu structure were shorter and were less variable than those for the broad structure. This is not surprising, however, since the overall trial times for the two menu structures were similar and the deep structure required an extra menu selection.

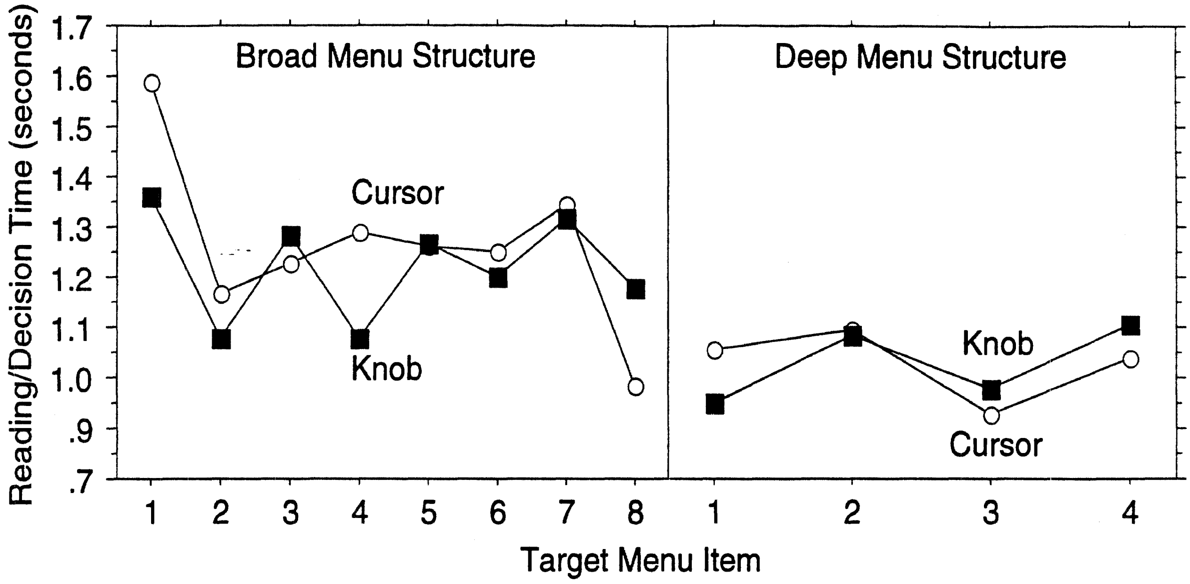


Figure 35. The effect of the menu item selected on the average reading/decision time.

The reading/decision times closely match predictions from the models of Miller (1980, 1981) and Landauer and Nachbar (1985) for simple selection times. Specifically, the reading/decision times presented here ranged from 0.98 to 1.59 seconds for the broad structure and from 0.93 to 1.11 seconds for the deep structure while the models predict a range of 1.08 to 2.08 seconds for an eight-item menu (which corresponds to the broad structure) and 0.76 to 1.84 seconds for a four-item menu (which corresponds to the deep structure). The models, however, included extra time for pushing a button while no time for any motor activity was included in the reading/decision times.

The final phase in building the GOMS models was to determine how the times for each step should be combined to estimate overall trial times. This was accomplished by generating a formula that took into account the reading/decision times for the particular menus involved and the number of down and enter keystrokes. The result was the following:

$$t(\text{total}) = \sum_{i=1}^{n(\text{menus})} [t(\text{mental})_i + [n(\text{downs})_i \times t(\text{down})] + t(\text{enter})]$$

where,

$t(\text{total})$ = the total time to complete the task
$n(\text{menus})$ = the number of menus encountered in completing the task
$t(\text{mental})_i$ = the time to read menu i and decide upon on action (the reading/decision time)
$n(\text{downs})_i$ = the number of downward scrolls to reach the desired item for menu i
$t(\text{down})$ = the time to press the down button or turn the knob one unit clockwise

This formula calculates the total time by taking the sum, for each menu, of the reading/decision time and the time for pressing buttons. The total time spent scrolling for any menu is the product of the number of down keystrokes and the time to execute one down keystroke.

The above formula was calculated for every possible combination of parameter values (i.e., number of menus, reading/decision time, number of downward scrolls, and time per scroll). Then, the actual trial times were averaged so that one mean was obtained corresponding to each estimated time. Finally, correlations were computed for each combination of menu structure and control to determine how well the model predicted the actual data (see Table 18).

Overall, the predictions of the GOMS model were within a quarter second of the actual values obtained from the experiment. The times were not accurate enough, however, to distinguish any differences between the menu structure or control. The correlation values ranged from 0.48 to 0.80, all significant at the 0.05 level.

Table 18. Average actual and predicted trial times and correlations by menu structure and control.

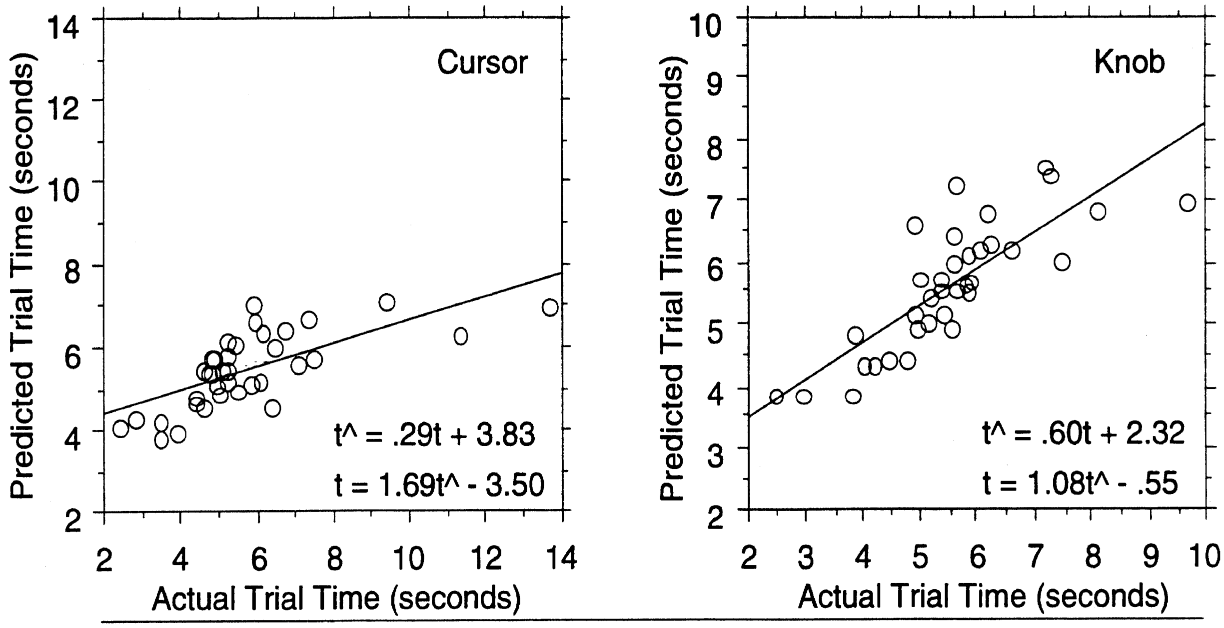
		Trial Time (seconds)			
Menu Structure	Control	Actual	Predicted	Residual	Correlation
Broad	Cursor	5.46	5.71	+.25	.69 (p < .001)
	Knob	5.61	5.53	-.08	.80 (p < .001)
Deep	Cursor	5.50	5.45	-.05	.68 (p < .001)
	Knob	5.76	5.51	-.25	.48 (p = .031)

From scatter plots of the predicted versus actual trial times (see Figures 36 and 37), it is apparent that a significant portion of the variability was not accounted for by the GOMS model. The GOMS estimates tended to overpredict short times and underpredict longer times. Consequently, the slopes of the regression equations were considerably less than unity while the Y-intercepts were greater than zero.

These departures of the model predictions may have occurred because this experiment did not fully satisfy one of the assumptions of GOMS modeling—that the operations performed by the user are well learned and routine. As indicated by the strong learning effect discussed in the two previous sections, subjects were still in the process of learning the task through most, if not all, of the experiment. The fact that the model tended to overestimate short times and underestimate long times could be a result of the use of means to estimate the down and enter keystroke times. Means may not have been the most typical values since relatively rare, extreme data points could have skewed the estimates. Also, drivers may have looked back to the road more frequently on long trials, an activity that would not be predicted by the model.

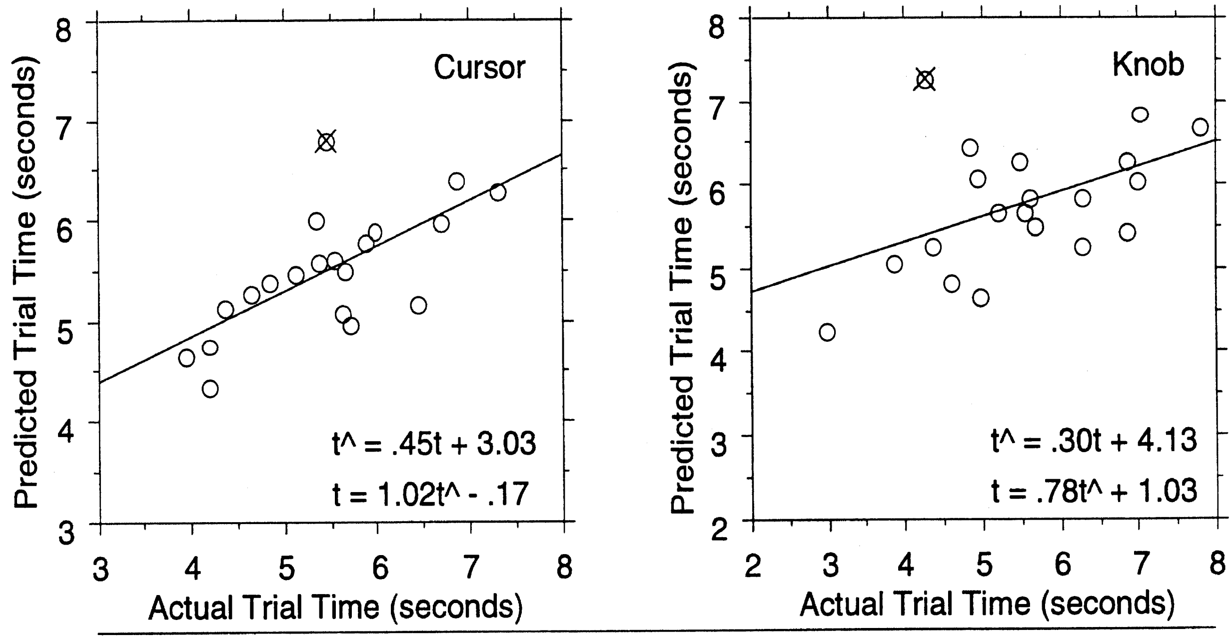
In addition, the model did not take into account such factors as how explicit the prompt questions were, how informative the various menu names were, or whether there were any words in common between menu titles and corresponding items (i.e., whether the menu items were overt or covert). In addition, the number of words or characters in each of the target items and the total number of words on the screen at any one time may have also been important factors. (The actual versus predicted times for each menu end node are provided in Appendix K to enable further exploration of these issues.)

For the deep menu structure, an outlier was identified for each of the two controls (denoted by *Xs* on the plots), in which the GOMS estimates were considerably larger than the actual times. Both of these outliers occurred for trials involving the last node of the hierarchy ("Off Road"), which involved scrolling to the bottom of each menu. Hence, the overestimation of the task time could have occurred because less data was available for this particular combination of scrolling. Also, it is possible that subjects looked ahead to the last item without scanning the middle ones for these particular trials. When the correlations were recalculated with these data points removed, the values improved to 0.78 for the cursor control and 0.69 for the knob.



t = actual trial time, t^ = predicted trial time

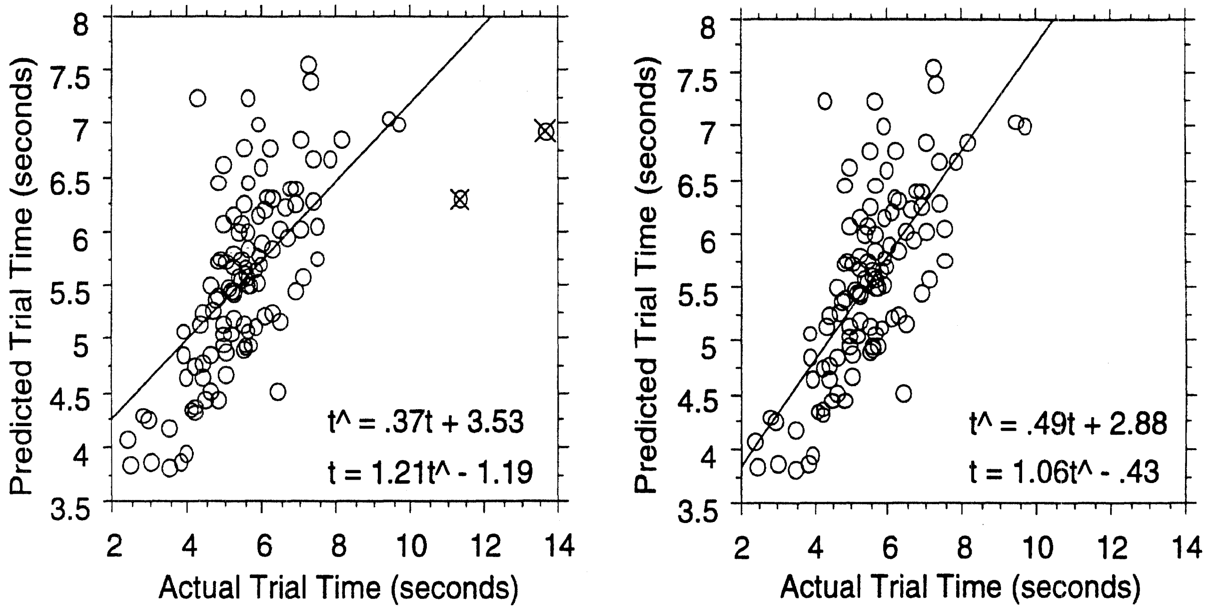
Figure 36. Scatter plot of predicted versus actual trial times for the broad menu structure.



t = actual trial time, t^ = predicted trial time

Figure 37. Scatter plot of predicted versus actual trial times for the deep menu structure.

Figure 38 shows regression plots and equations of GOMS estimates versus actual trial times for all the data combined. With all data points included, the correlation was 0.67. When outliers (denoted by *Xs*) were removed, the correlation increased to 0.72. As the formula in the left graph indicates, a good approximation of the actual time (in seconds) is to multiply the GOMS estimate by 1.2 and then subtract 1.2.



t = actual trial time, t^ = predicted trial time

Figure 38. Scatter plot of predicted versus actual trial times for all the data combined (left) and with two outliers removed (right).

Survey results

The data from the acceptability survey (see Appendix I) gives a qualitative indication of how participants perceived the menu interface. Table 19 provides the results for question 1, which requested an overall impression of the system. Sixty-three percent of the subjects had positive comments, although several of those with positive comments also had negative ones as well. Five of the 24 participants had purely negative overall impressions. Although opinions on the features of the system were not directly solicited, four subjects expressed a preference for the deep menus while none did so for the broad.

Table 19. Responses to Question 1—"What is your overall impression of the computer menu system you just used?"

	Comment		Expressed Preference*			
Subject	Positive	Negative	C	K	B	D
1	"Good overall"	"Sometimes confusing"				
2	"Interesting"					
3	"Can be useful"	"Somewhat hazardous"				
4		"Not intuitive"	√			
5	"Easy"					√
6	"Nice font"					√
7	"Promising"			√		
8		"I do not like looking away from the road"				
9		"Intense concentration"				
10	"Very informative"					
11	"Pretty good"					
12						
13	"Great idea"	"Hope it doesn't distract people"				
14						√
15		"It was awkward"				
16	"Good"	"A concern would be electronic failure"				
17						
18	"Very easy to use and understand"					
19	"I liked it"					
20	"Very effective"	"Not familiar with some words"				
21	"Thought it was fun"					√
22		"Horrible!"				
23		"Too much reading and eyes leaving the road"				
24	"Very interesting"					
Count	15	11	1	1	0	4
Percent	63	46	4	4	0	17

* C = cursor control, K = knob control, B = broad menus, and D = deep menus

Table 20 summarizes the rankings for question 2. Two-thirds of subjects thought "a system that understands voice commands" would be the best and a system similar to the one used in the study would be second best, while 79 percent of subjects thought "a system that uses knobs, buttons, and switches, but consists of a larger number of them placed closer together than in today's cars" would be the worst solution. Fifteen of the 24 subjects (62.5 percent) ranked the choices in the order voice first (b), then menus (a), and then controls (c).

Table 20. Responses to question 2—"How would you rank the following alternatives?"

Alternative	Count (Percent) of Subjects		
	1st Choice	2nd Choice	3rd Choice
a. Computer menu system	7 (29)	16 (67)	1 (4)
b. Voice recognition system	16 (67)	4 (17)	4 (17)
c. Larger number of controls	1 (4)	4 (17)	19 (79)

The responses for question 3, which requested opinions on the advantages of the system, were divided into six categories (see Table 21). Most subjects cited space efficiency, having all the controls in one location, ease of finding the control, or minimal dash clutter as advantages. Some also noted that such a system allows for a number of features that are not presently available. Finally, six subjects provided responses that did not fall into any of the categories, while four subjects expressed that there were no advantages whatsoever.

Table 21. Responses to question 3—"What do you feel are the advantages of using a computer menu system?"

Expressed Advantage	Count (Percent) of Subjects
Space efficiency	3 (9)
All controls in one place	4 (12)
Easier to find control	3 (9)
Less dash clutter	4 (12)
More features	4 (12)
None	4 (12)

Table 22 shows the system disadvantages. By far the most common complaint (71 percent of participants) was that the system was distracting. Phrases such as "distracting from my driving" and "reading menus instead of driving" were very typical. Subjects also thought the system was confusing and difficult to learn. In agreement with question 1, several subjects expressed disapproval of the broad menus.

Table 22. Responses to question 4—"What do you feel are the disadvantages of using a computer menu system?"

Expressed Disadvantage	Count (Percent) of Subjects
Distracting	17 (71)
Confusing	2 (8)
Difficult to learn	4 (17)
Disliked broad menus	5 (21)

Eye fixations

To recap, observations about glancing behavior were made during the experiment for all subjects. Specifically, for each subject, the minimum and maximum number of glances at the display per menu was recorded for approximately four randomly selected trials for each block, a total of about 1152 trials. Since the intention was only to gain a sense of how often drivers glanced at the display, no formal analyses were performed. However, three observations can be made. First, the minimum number of glances per menu ranged from one to two with a mean of just slightly above one. Second, the maximum number of glances per menu ranged from one to seven with a mean of approximately 2.4.

Third, the limited data suggests that the broad menu structure may have led to more glances per menu than the deep structure, especially among older subjects (see Figure 39). This can easily be explained by the greater content of the broad menus. This does not suggest, however, that the number of glances per trial was higher.

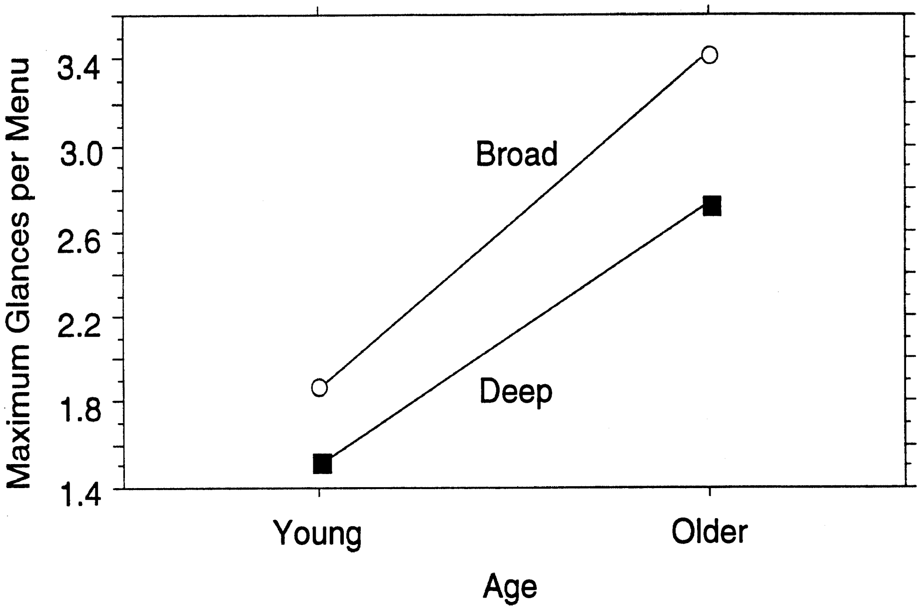


Figure 39. The effect of menu structure and age on the maximum number of glances per menu.

Initial subject reactions

Several interesting observations were made regarding participants' initial reactions to the in-vehicle menu system by the experimenter. First, it was apparent that the system was much more difficult for the older participants to learn than the young ones. Many of them required several trials with experimenter assistance before they could perform the task on their own. Additionally, several of them had a great deal of difficulty keeping the simulated vehicle on the road during the first two practice blocks. In a few cases, they resorted to stopping or slowing down considerably until they understood the system well enough to handle both tasks at once. In contrast, nearly all of the young subjects understood how the system worked almost immediately and asked very few questions, although some of them had minor difficulties maintaining their course at first.

With regard to specific attributes of the interface, the greatest initial problem for most subjects was learning which menus to choose for each requested item, since one or two selections were required before the desired item could be found. The young subjects, however, tended to learn very quickly how the prompted item was categorized, while many of the older subjects took much longer, and in some cases, never found certain items. Few subjects had difficulty understanding how the controls worked, although some of the older subjects needed reminders during their first few practice trials. Finally, no subjects appeared to have any trouble with display legibility, although many felt the broad menus were daunting due to their large amount of text.

The final observation is that initial impressions of the older participants were often quite pessimistic. Many felt they would not be able to learn the system after some early difficulties, and some simply expressed dislike for it. The young subjects, on the other hand, tended to have more positive initial impressions.

CONCLUSIONS

What are typical selection times and error rates for novice users of a simulated in-vehicle menu system?

The best estimates of selection times and error rates are from the main experiment. The mean time was just 9.1 seconds, with times ranging from 1.3 to 87.3 seconds. For ideal trial (no selection errors, no backtracking), the mean was 7.6 seconds with a maximum of 49.1 seconds. Readers should keep in mind that this time is for a 64-node interface, and that changing the number of items would change the times. For interfaces now under consideration, a 64-node interface is reasonable.

The mean error rate was 7.8 percent. There were three types of errors: (1) selecting an incorrect item believed to be correct, (2) selecting an incorrect item after being unable to find the correct one, or (3) selecting an item inadvertently. Data on the frequency of each type of error were not collected. Error rates for older drivers were unacceptably high, 13 percent, though these error rates would probably be lower for menu structures optimized for usability (rather than experimental efficacy).

Can selection times be reliably predicted using a GOMS model?

Correlations of the GOMS predictions with the actual times varied between 0.48 for deep menus using the knob and 0.80 for broad menus using the knob. In considering these values, one must keep in mind that the "actual" times reported are estimates of the actual times from a small sample of subjects for a limited number of trials, not the true times. Given the sample size, these estimates are quite good.

Generally, the GOMS model overpredicted short trial times and underpredicted longer times, an error correctable using regression adjustments. The worst case was for the cursor control and the broad menus where the actual time was approximately 14 seconds and the uncorrected prediction was approximately seven seconds.

The GOMS model predictions were imperfect in part because the model assumes subjects are performing a routine cognitive task, whereas in the experiment subjects were still learning how to use the menu system. In fact, due to the nature of driving, some drivers many have never known the full interface well enough for its operation to be a routine task. Furthermore, because driving involves time sharing, subjects needed to periodically abandon the in-vehicle task to attend to the road.

While not perfect, these results suggest that GOMS models can be used to predict driver performance with hierarchical menu systems prior to building any prototypes. Additional work is needed, however, to relate driving workload to total task times and to further determine the sources of differences between actual and estimated times.

What is the effect of driver age and sex on selection times and error rates?

The effects of driver sex and, especially, age on performance were substantial. Consistent with previous studies, trial times for older drivers were almost double those of younger drivers (12.1 versus 6.2 seconds). Error rates differed by more than a factor of five (13.3 percent for older drivers and 2.3 percent for younger drivers). These differences probably reflect both true age differences and the fact that younger subjects had far more computer experience. Finally, older participants engaged in more backtracking than young ones. This could be an indication that the menu structures were too extensive for some older subjects to learn fully.

In terms of overall gender differences, women took about 30 percent longer to respond than men (10.4 versus 7.9 seconds) and made about 50 percent more errors (9.3 versus 6.2 percent).

How do selection times and error rates vary as a function of menu structure, control type, and the location of the control and display?

Although the effect of menu structure on both trial times and error rates was not statistically significant, deep menus took about two percent longer to retrieve and had error rates that were under three percent greater. Also, the broad menus yielded more scrolling errors while the deep structure caused more backtracking errors. Hence, there was no practical difference due to menu structure and therefore no clear resolution to the issue of depth versus breadth.

With regard to control, the pilot test, which involved young subjects only, showed that trial times for the cursor control were about 14 percent shorter than those for the number pad (4.2 versus 4.8 seconds). There were too few errors to evaluate the effect on accuracy. In the main experiment, the knob control was about nine percent faster than the cursor control and led to about three percent fewer errors. Hence, the knob is the preferred control from the performance perspective. Furthermore, knob controls where the selection (enter) button is integrated with the knob itself may offer an even greater advantage.

Finally, the control and display location did not significantly affect performance, with mean times varying by only seven percent between the three configurations and error rates by somewhat more, 16 percent. If forced to choose, the recommendation is to place the display high (close to the driver's line of sight) and the control low (within easy reach).

To what extent do selection times and error rates decline with practice?

The effect of practice was statistically significant and considerable. Mean trial times dropped from 7.5 to 5.8 seconds (29 percent) from the first to the twelfth block (192 total responses). Times for older drivers dropped from 15 seconds to 9.5, a decrease of 58 percent. Due to the limited sample size, block-to-block variations in error rate were on the order of several percent, but the drop was from about 11 percent to seven percent, a considerable difference.

What predictions can be made about the safety of in-vehicle menu interfaces?

Three direct measures of safety were examined: lane excursions per block, standard deviation lane position, and standard deviation speed. Clearly, the more often the driver departs from the lane, weaves within the lane, and varies his or her driving speed, the greater the risk of an accident. Interestingly, none of the factors of interest (age, gender, menu structure, control, configuration, or block) had any effect on standard deviation lane position. Similarly, the effect of standard deviation speed was not significant for any factor except control, where speeds were more variable for the cursor control (7.9 miles per hour versus 7.4 for the knob). With regard to lane excursions, older drivers committed significantly more than younger drivers (8.0 versus 2.9 per block). Also, there were significantly more lane excursions for broad menus than for deep (5.7 versus 5.0), and more for the first block than for the last (6.2 versus 4.5). The values reported for standard deviation lane position and speed (1.25 feet and 7.4 miles per hour) are approximately triple the values reported in the literature for driving alone. This indicates that the menu task was a significant distraction.

It seems reasonable that the 14 percent increase in the number of lane excursions for broad menus was due to the increased visual demands. The data show that more glances per menu were required for broad menus than for deep menus (1.9 versus 1.5 for young drivers, 2.8 versus 3.5 for older drivers), about a 26 percent increase. Although this does not imply that performing the entire task required more glances since the broad structure involved fewer menus, it does suggest that the deep menus may have better facilitated attention sharing between the task and the road. There was insufficient data to allow for an analysis comparing eye-fixation frequencies to lane excursions.

Overall, the performance and eye-fixation data suggest that caution is warranted in implementing an in-vehicle menu system, and if one is to be deployed, deep menus are advised over broad.

How acceptable is the idea of an in-vehicle menu interface to younger and older drivers?

From the limited sample, a representative set of impressions reflecting all drivers is difficult to obtain. Of those responding, positive post-test comments outnumber negative ones (15 to 11), but that could be due partly to respondents being polite or trying to balance negative comments with positive ones. Drivers recognized that the interface could save space, but expressed concern that it would be distracting. Several drivers felt the broad menus were particularly distracting or went out of their way to note a preference for the deep menus even though they were not asked to comment about them.

As evidence of the driver distraction noted in the surveys, the mean number of lane excursions per block was 5.7, a rather large number for 16 trials occurring within a 10 minute drive. Unfortunately, normative data for driving alone or driving while using other interface types does not exist, in particular for the test road used in this experiment. However, the experimenter was fairly certain that most lane excursions were caused by the attentional demands of the menu interface, not because the driving task itself was challenging.

The experimenter also observed that older drivers experienced challenges in learning the menu structures and integrating the menu task with driving, while younger drivers adapted much more easily. Some older drivers required direct experimenter assistance to complete the first few trials and others experienced great difficulty in keeping the vehicle on the road during the first two practice blocks. Some older drivers slowed down considerably (or even stopped) to focus on the menu selection task. Younger subjects, on the other hand, were able to handle both driving and the menu task together within a few trials, except for a few who had initial difficulty maintaining their course.

How many menu end nodes should be explored per block?

The pilot experiment involved 32 trials (i.e., half of the available end nodes) per block. Hence, for each block, subjects selected either two items from each final menu for the deep (4 X 4 X 4) structure or four items from each final menu for the broad (8 X 8) structure. An ANOVA of the data for the four subjects suggested that there was no statistically significant difference between analyses of 1/2 and 1/4 of the end nodes (16 and 32 trials per block, respectively). Hence, the design for the main experiment only called for one end node from each final menu of the deep structure to be tested per block. This guideline may be applicable to future studies involving menus.

Concluding remarks

As a result of this study, the following three design guidelines emerged:

1. Knob controls are preferable to cursor-type controls and number pads.
2. Menu depth versus breadth had little impact on selection time and errors, but a deep menu structure is advisable since the deep menus led to fewer lane excursions and several subjects expressed a preference for them.
3. Locating the display high on the instrument panel and the control low is recommended, but differences between the various locations were not statistically significant.

This study suggests several topics for future research. First, baseline data on time, errors, and lane excursions for driving alone and using dedicated instrument-panel controls would enhance the value of these findings. In addition, the performance of other likely implementations of in-vehicle menu interfaces, such as mounting the control on the steering wheel or using a head-up display (HUD), would provide an important perspective. Finally, further exploration of the impact of alternative menu names would be valuable in helping to build highly usable driver interfaces.

REFERENCES

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

Fitts, P.M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement, Journal of Experimental Psychology, 47, 381-391.

Green, P. (1996). Driver Interfaces for Vehicles of the Future, Inside Automotives, November/December, 3(6), 33-36.

Green, P. (1995). Suggested Procedures and Acceptance Limits for Assessing the Safety and Ease of Use of Driver Information Systems (Technical Report FHWA-RD-94-089), McLean, VA: U. S. Department of Transportation, Federal Highway Administration.

Green, P. (1979). Automobile Multifunction Stalk Controls: Literature, Hardware and Human Factors Review (Technical Report UM-HSRI-79-78), Ann Arbor, MI: University of Michigan Highway Safety Research Institute.

Green, P., Serafin, C., Williams, M., and Paelke, G. (1991). What Functions and Features Should Be in Driver Information Systems of the Year 2000? Vehicle Navigation and Information Systems Conference (VNIS'91), (SAE paper 912792), Warrendale, PA: Society of Automotive Engineers, 483-498.

Heuchert, S. (1995). Eyes Forward - an Ergonomic Solution to Driver Information Overload (paper 9531183), The Eighth International Pacific Conference on Automotive Engineering (IPC-8), Tokyo, Japan: Society of Automotive Engineers of Japan, 169-174.

Hyman, R. (1953). Stimulus Information as a Determinant of Reaction Time, Journal of Experimental Psychology, 45, 188-196.

Katz, S., Fleming, J., Green, P., Hunter, D.R., and Damouth, D. (1997). On-the-Road Human Factors Evaluation of the Ali-Scout Navigation System (Technical Report UMTRI-96-32), Ann Arbor, MI: University of Michigan Transportation Research Institute.

Landauer, T.K. and Nachbar, D.W. (1985). Selection From Alphabetic and Numeric Menu Trees Using A Touch Screen: Breadth, Depth, and Width, CHI'85 Proceedings, New York, NY: Association of Computing Machinery, 73-78.

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

Manes, D., Green, P., and Hunter, D. (1997). Prediction of Destination Entry and Retrieval Times Using GOMS (Technical Report UMTRI-96-37), Ann Arbor, MI: University of Michigan Transportation Research Institute.

Miller, D.P. (1980). Factors Affecting Item Acquisition Performance in Hierarchical Systems: Depth vs. Breadth (Ph.D. dissertation), Columbus, OH: Ohio State University.

Miller, D.P. (1981). The Depth/Breadth Tradeoff in Hierarchical Computer Menus. Proceedings of the Human Factors Society 25th Annual Meeting, Santa Monica, CA: Human Factors and Ergonomics Society, 296-300.

Musseler, J. (1994). Using Predictors to Partition Menu Selection Times, Behavior and Information Technology, 13(6), 362-372.

Norman, K.L. (1991). The Psychology of Menu Selection, Norwood, NJ: Ablex.

Paelke, G.M. (1993). A Comparison of Route Guidance Destination Entry Methods, Proceedings of the Human Factors and Ergonomics Society, 37th Annual Meeting, Santa Monica, CA: 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: University of Michigan Transportation Research Institute.

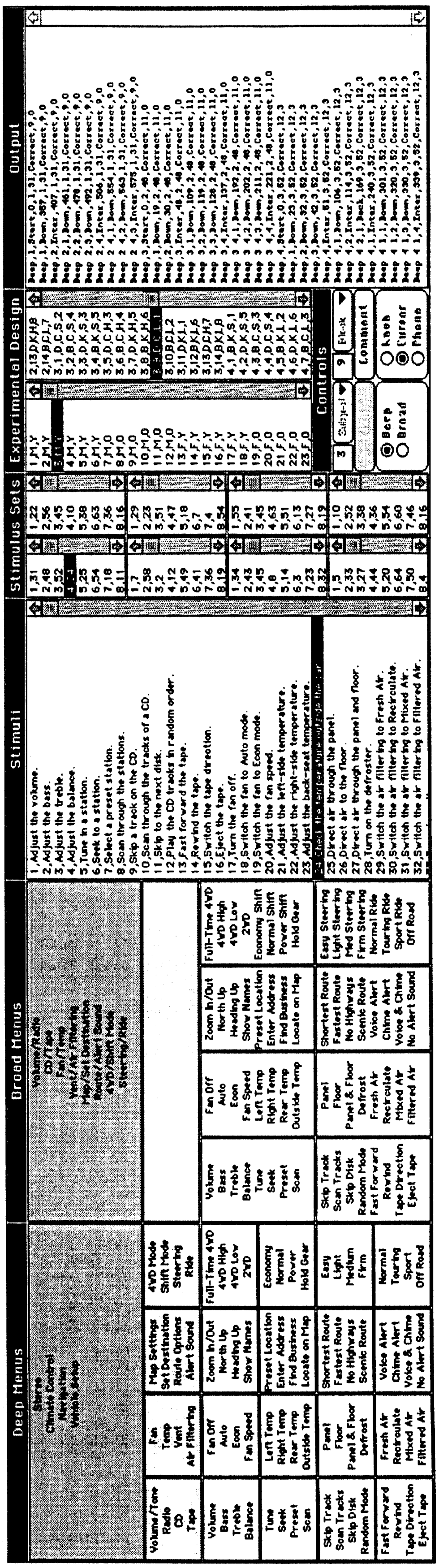
Reed, M. P. and Green, P. (1995). Validation of a Low-Cost Driving Simulator Using a Telephone Dialing Task (Technical Report UMTRI-95-19), Ann Arbor, MI: University of Michigan Transportation Research Institute.

Shinar, D., Stern, H.I., Dubis, G., and Ingram, D. (1985). The Relative Effectiveness of Alternative Selection Strategies in Menu Driven Computer Programs, Proceedings of the Human Factors Society, 29th Annual Meeting, Santa Monica, CA: Human Factors Society, 645-649.

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: University of Michigan Transportation Research Institute.

Sumie, M., Li, C., and Green, P. (1997). Usability of Menu-Based Interfaces for Motor Vehicle Secondary Functions (Technical Report UMTRI-97-19), Ann Arbor, MI: University of Michigan Transportation Research Institute (in preparation).

APPENDIX A - SCREEN SHOT OF THE MENUPLAYER CONTROL PANEL



Stimuli

1,Adjust the volume.
2,Adjust the bass.
3,Adjust the treble.
4,Adjust the balance.
5,Tune in a station.
6,Seek to a station.
7,Select a preset station.
8,Scan through the stations.
9,Skip a track on the CD.
10,Scan through the tracks of a CD.
11,Skip to the next disk.
12,Play the CD tracks in random order.
13,Fast forward the tape.
14,Rewind the tape.
15,Switch the tape direction.
16,Eject the tape.
17,Turn the fan off.
18,Switch the fan to Auto mode.
19,Switch the fan to Econ mode.
20,Adjust the fan speed.
21,Adjust the left-side temperature.
22,Adjust the right-side temperature.
23,Adjust the back-seat temperature.
24,Check the temperature outside the car
25,Direct air through the panel.
26,Direct air to the floor.
27,Direct air through the panel and floor.
28,Turn on the defroster.
29,Switch the air filtering to Fresh Air.
30,Switch the air filtering to Recirculate.
31,Switch the air filtering to Mixed Air.
32,Switch the air filtering to Filtered Air.

Stimulus Sets

```
1,31   1,22
2,48   2,56
3,52   3,45
4,34   4,10
5,25   5,38
6,54   6,63
7,18   7,36
8,11   8,16

1,7    1,29
2,58   2,23
3,2    3,51
4,12   4,47
5,49   5,18
6,41   6,7
7,36   7,4
8,19   8,54

1,34   1,55
2,43   2,41
3,45   3,45
4,8    4,63
5,14   5,51
6,3    6,13
7,23   7,27
8,32   8,19

1,5    1,10
2,33   2,52
3,27   3,38
4,44   4,36
5,20   5,54
6,64   6,60
7,50   7,46
8,4    8,16
```

Experimental Design

```
1,M,Y    2,13,D,K,H,8
2,M,Y    2,14,B,C,L,7
3,M,Y    3,1,D,C,S,2
4,M,Y    3,2,B,C,S,4
5,M,Y    3,3,D,K,S,3
6,M,Y    3,4,B,K,S,5
7,M,O    3,5,D,C,H,3
8,M,O    3,6,B,C,H,4
9,M,O    3,7,D,K,H,5
10,M,O   3,8,B,K,H,6
11,M,O   3,9,D,C,L,1
12,M,O   3,10,B,C,L,2
13,F,Y   3,11,D,K,L,1
14,F,Y   3,12,B,K,L,6
15,F,Y   3,13,D,C,H,7
16,F,Y   3,14,B,K,L,8
17,F,Y   4,1,B,K,S,1
18,F,Y   4,2,D,K,S,5
19,F,O   4,3,B,C,S,3
20,F,O   4,4,D,C,S,4
21,F,O   4,5,B,K,L,2
22,F,O   4,6,D,K,L,6
23,F,O   4,7,B,C,L,3
```

Output

```
Deep,1,Start,0,1,31,Correct,9,0
Deep,1,Down,367,1,31,Correct,9,0
Deep,2,Inter,407,1,31,Correct,9,0
Deep 2,1,Down,461,1,31,Correct,9,0
Deep 2,2,Down,478,1,31,Correct,9,0
Deep 2,3,Down,492,1,31,Correct,9,0
Deep 2,4,Inter,506,1,31,Correct,9,0
Deep 2 4,1,Down,554,1,31,Correct,9,0
Deep 2 4,2,Down,563,1,31,Correct,9,0
Deep 2 4,3,Inter,575,1,31,Correct,9,0
Deep,3,Start,0,2,48,Correct,11,0
Deep,1,Down,19,2,48,Correct,11,0
Deep,2,Down,30,2,48,Correct,11,0
Deep,3,Inter,48,2,48,Correct,11,0
Deep 3,1,Down,109,2,48,Correct,11,0
Deep 3,2,Down,119,2,48,Correct,11,0
Deep 3,3,Down,128,2,48,Correct,11,0
Deep 3,4,Inter,137,2,48,Correct,11,0
Deep 3 4,1,Down,192,2,48,Correct,11,0
Deep 3 4,2,Down,202,2,48,Correct,11,0
Deep 3 4,3,Down,211,2,48,Correct,11,0
Deep 3 4,4,Inter,221,2,48,Correct,11,0
Deep,4,Start,0,3,52,Correct,12,3
Deep,1,Down,23,3,52,Correct,12,3
Deep,2,Down,32,3,52,Correct,12,3
Deep,3,Down,42,3,52,Correct,12,3
Deep,4,Inter,51,3,52,Correct,12,3
Deep 4,1,Down,104,3,52,Correct,12,3
Deep 4,2,Inter,114,3,52,Correct,12,3
Deep 4 2,1,Back,169,3,52,Correct,12,3
Deep 4,1,Inter,240,3,52,Correct,12,3
Deep 4 1,1,Down,301,3,52,Correct,12,3
Deep 4 1,2,Down,320,3,52,Correct,12,3
Deep 4 1,3,Down,330,3,52,Correct,12,3
Deep 4 1,4,Inter,339,3,52,Correct,12,3
```

APPENDIX B - CONTROL AND DISPLAY LOCATIONS

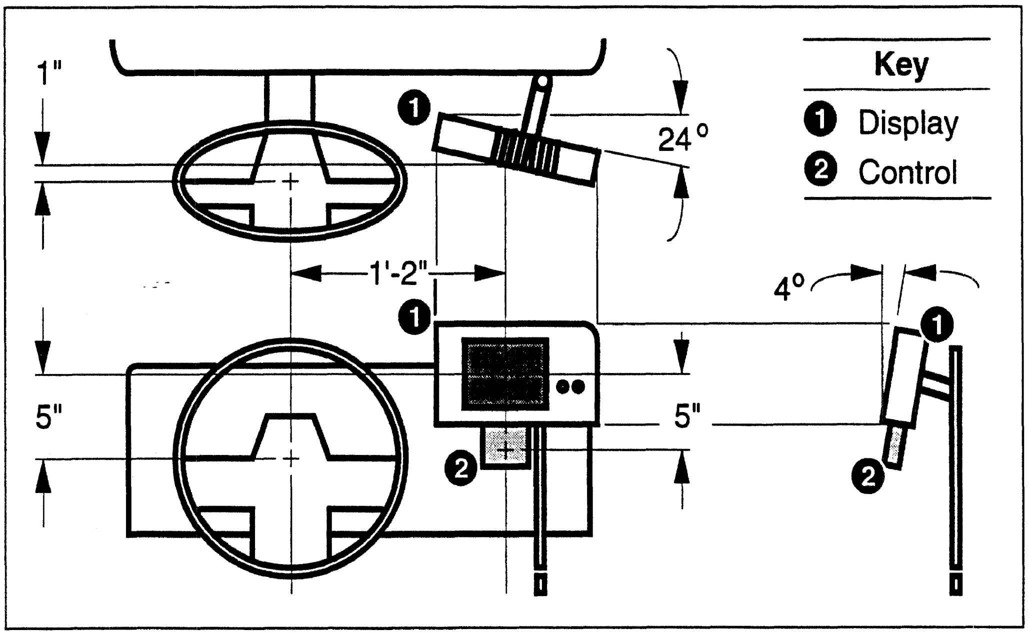


Figure 40. The control and display locations for the high configuration.

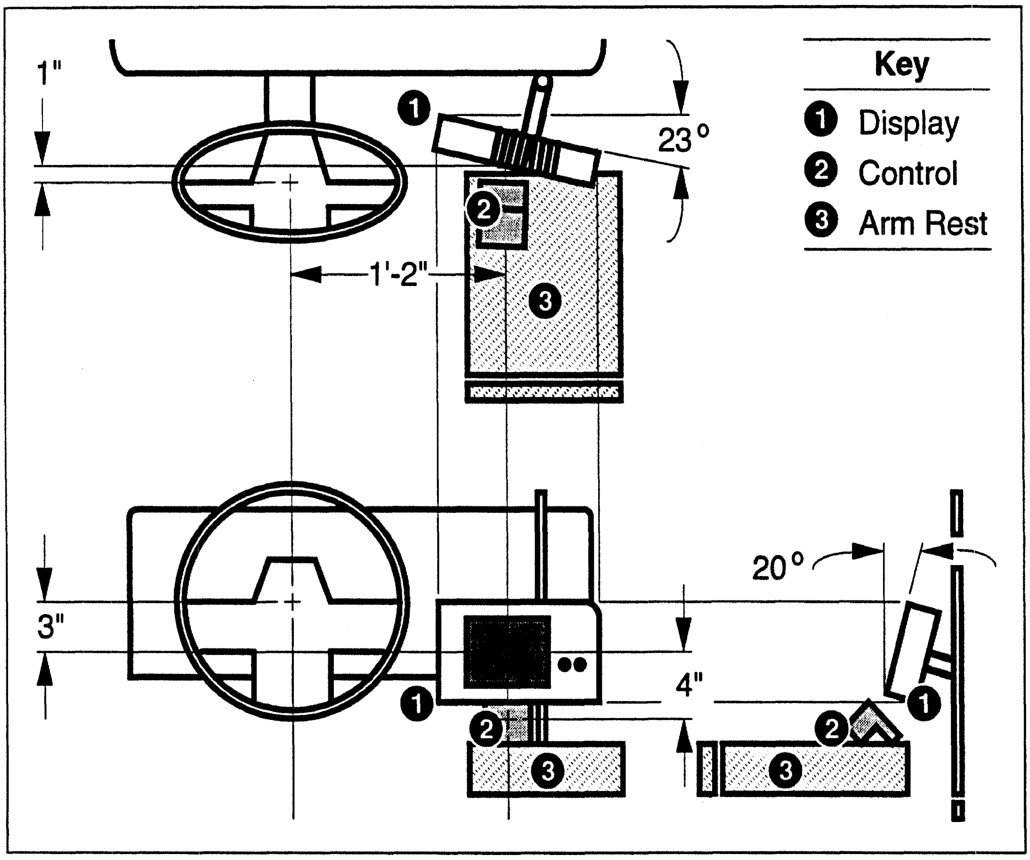


Figure 41. The control and display locations for the low configuration.

APPENDIX C - TASK PROMPTS

#	How would you...
1	adjust the volume?
2	adjust the bass?
3	adjust the treble?
4	adjust the balance?
5	tune in a station?
6	seek to a station?
7	select a preset station?
8	scan through the stations?
9	skip a track on the CD?
10	scan through the tracks of a CD?
11	skip to the next disk?
12	play the CD tracks in random order?
13	fast forward the tape?
14	rewind the tape?
15	switch the tape direction?
16	eject the tape?
17	turn the fan off?
18	switch the fan to Auto mode?
19	switch the fan to Econ mode?
20	adjust the fan speed?
21	adjust the left-side temperature?
22	adjust the right-side temperature?
23	adjust the back-seat temperature?
24	check the temperature outside the car?
25	direct air through the panel?
26	direct air to the floor?
27	direct air through the panel and floor?
28	turn on the defroster?
29	switch the air filtering to Fresh Air?
30	switch the air filtering to Recirculate?
31	switch the air filtering to Mixed Air?
32	switch the air filtering to Filtered Air?
33	zoom in on the map?
34	have the map displayed north-up?
35	have the map displayed heading-up?
36	have the map show street names?
37	set a destination using a preset location?
38	set a destination by entering an address?
39	set a destination by searching businesses?
40	set a destination by locating it on the map?
41	obtain the shortest route?
42	obtain the fastest route?
43	obtain a route with no highways?
44	obtain a scenic route?
45	receive navigation alerts by voice?
46	receive navigation alerts by chime?
47	receive navigation alerts by voice and chime?
48	turn all navigation alerts off?
49	switch to Full-Time 4WD?
50	switch to 4WD High?
51	switch to 4WD Low?
52	switch to 2WD?
53	set the shift mode to Economy?
54	set the shift mode to Normal?
55	set the shift mode to Power?
56	set the shift mode to Hold Gear?
57	make the steering effort easy?
58	make the steering effort light?
59	make the steering effort medium?
60	make the steering effort firm?
61	set the ride to Normal?
62	set the ride to Touring?
63	set the ride to Sport?
64	set the ride to Off-Road?

APPENDIX D - PROMPT SETS

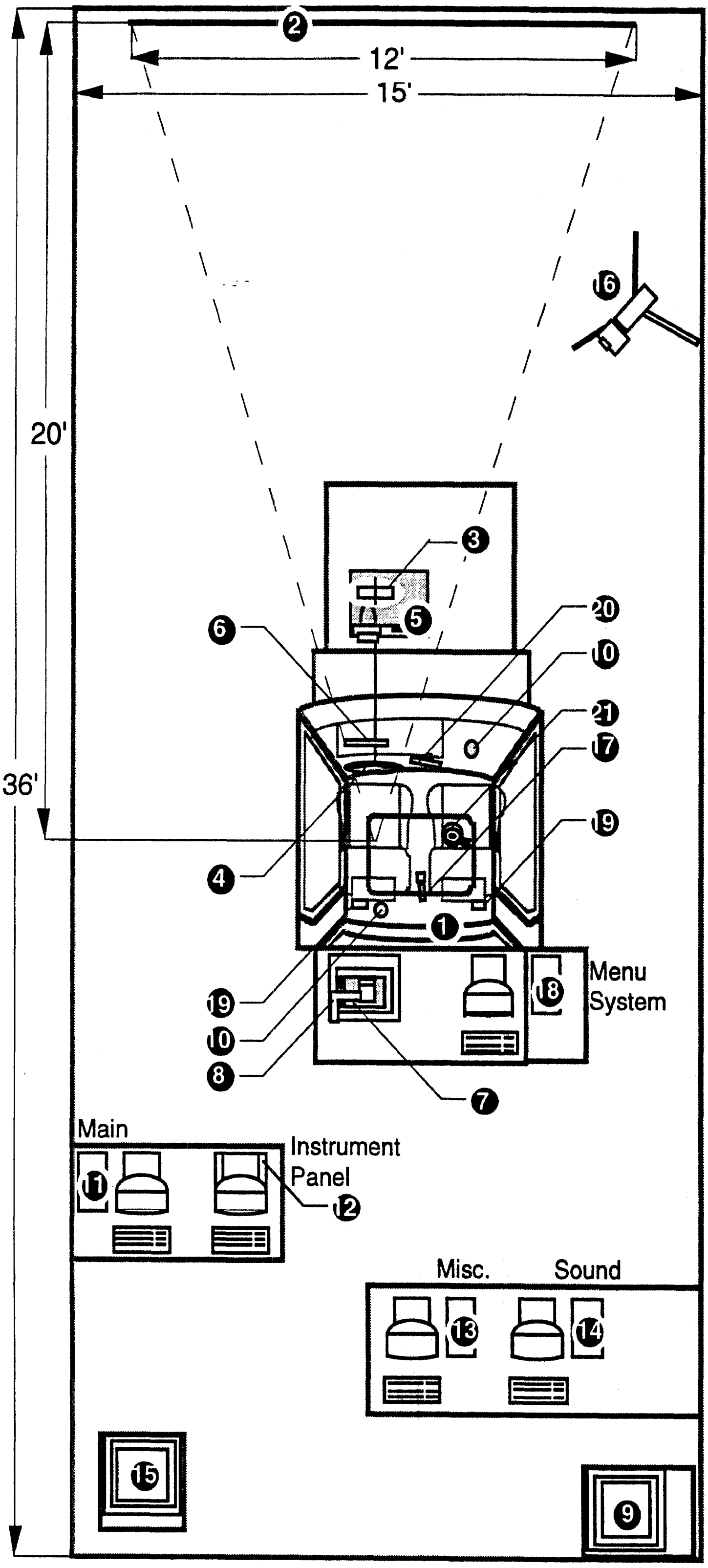
Table 23. Prompt sets used in the pilot experiment.

#	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
1	24	5	4	41	36	47
2	51	8	50	24	20	49
3	43	10	58	54	8	43
4	9	34	48	33	4	25
5	39	62	21	45	34	57
6	29	16	31	44	53	38
7	1	30	36	50	60	48
8	56	21	30	46	50	36
9	32	19	44	55	61	22
10	44	63	26	38	13	62
11	34	48	10	26	27	19
12	53	58	25	6	28	6
13	58	46	49	21	46	60
14	14	39	7	40	51	61
15	15	11	18	63	22	4
16	38	22	63	64	42	3
17	62	40	16	14	23	55
18	23	14	54	52	17	24
19	57	60	56	57	7	27
20	20	20	35	12	59	11
21	45	53	39	2	15	18
22	18	3	6	32	47	42
23	52	25	59	3	40	37
24	26	54	15	27	37	16
25	2	1	23	35	29	9
26	47	28	17	17	9	5
27	64	52	37	13	2	51
28	33	35	61	11	55	56
29	12	29	1	8	31	13
30	7	43	12	19	64	32
31	5	49	45	59	41	33
32	28	41	42	31	10	30

Table 24. Prompt sets used in the main experiment.

#	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
1	31	7	34	5	22	29
2	48	58	43	33	56	23
3	52	2	45	27	45	51
4	24	12	8	44	10	47
5	25	49	14	20	38	18
6	54	41	3	64	63	7
7	18	36	23	50	36	4
8	11	19	32	4	16	54
9	37	13	51	15	50	28
10	42	46	17	10	43	33
11	16	56	62	38	57	40
12	1	22	53	57	27	60
13	6	39	40	30	5	42
14	61	28	60	47	32	14
15	35	63	9	21	17	61
16	59	29	26	55	3	9

APPENDIX E. DRIVER INTERFACE RESEARCH SIMULATOR



① 1985 Chrysler Laser mockup with simulated hood
② 8'X10' projection screen with 3M hi-white encapsulated reflective sheeting
③ PMI Motion Technologies ServoDisk DC motor (model 00-01602-002 type U16M4) with Copley Controls Corp. controller (model 413) and power supply (model 645)
④ 3-spoke steering wheel
⑤ Sharp color LCD projection system (model XG-E850U) for instrument panel
⑥ 4"X13" plexiglas screen
⑦ Sharp computer projection panel (model QA-1650)
⑧ 3M overhead projector (model 9550)
⑨ Audio rack
⑩ Aura bass shakers (model AST-1B-4) for vibration
⑪ Power Macintosh 9500/200
⑫ Power Macintosh 7100/80AV
⑬ Power Macintosh 8500/120
⑭ Macintosh Quadra 840AV
⑮ Video rack
⑯ Panasonic low level light camera (model WV-BP510)
⑰ Panasonic "lipstick" camera (model GP-KS152)
⑱ Power Macintosh 8100/100
⑲ Sony speaker system (model SR6-48)
⑳ Cascade Technology Corp. DiscoveryMATE 6" LCD display (model R65)
㉑ Lamp with 15-watt bulb to illuminate interior

Figure 42. Schematic of the Driver Interface Research Simulator.

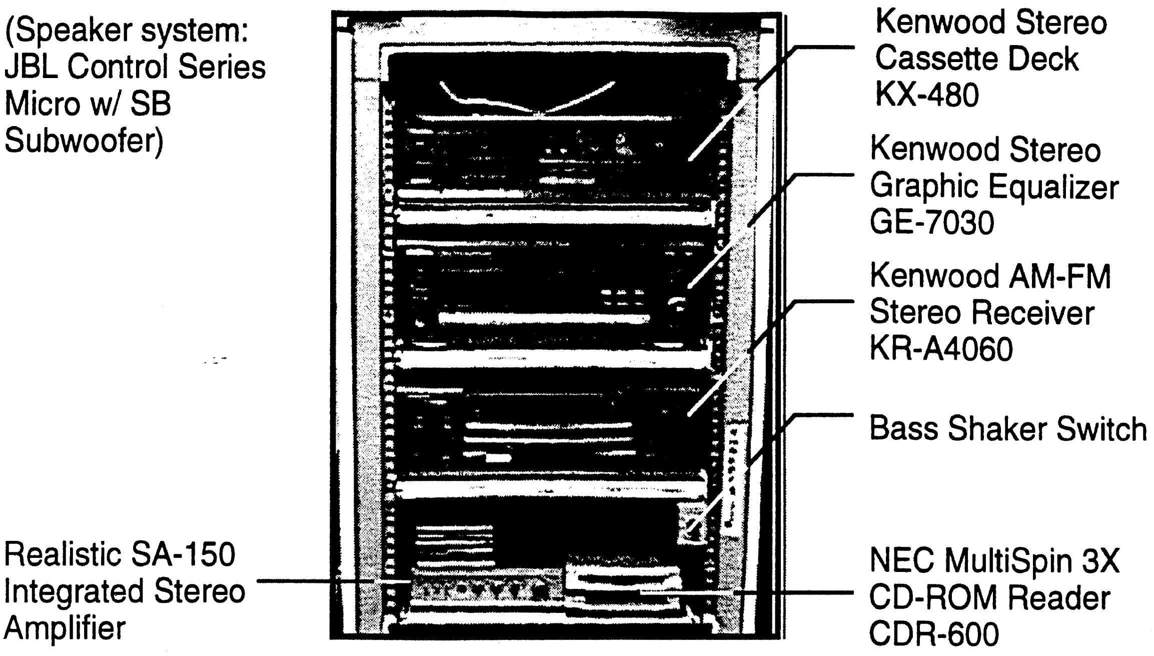


Figure 43. The audio rack.

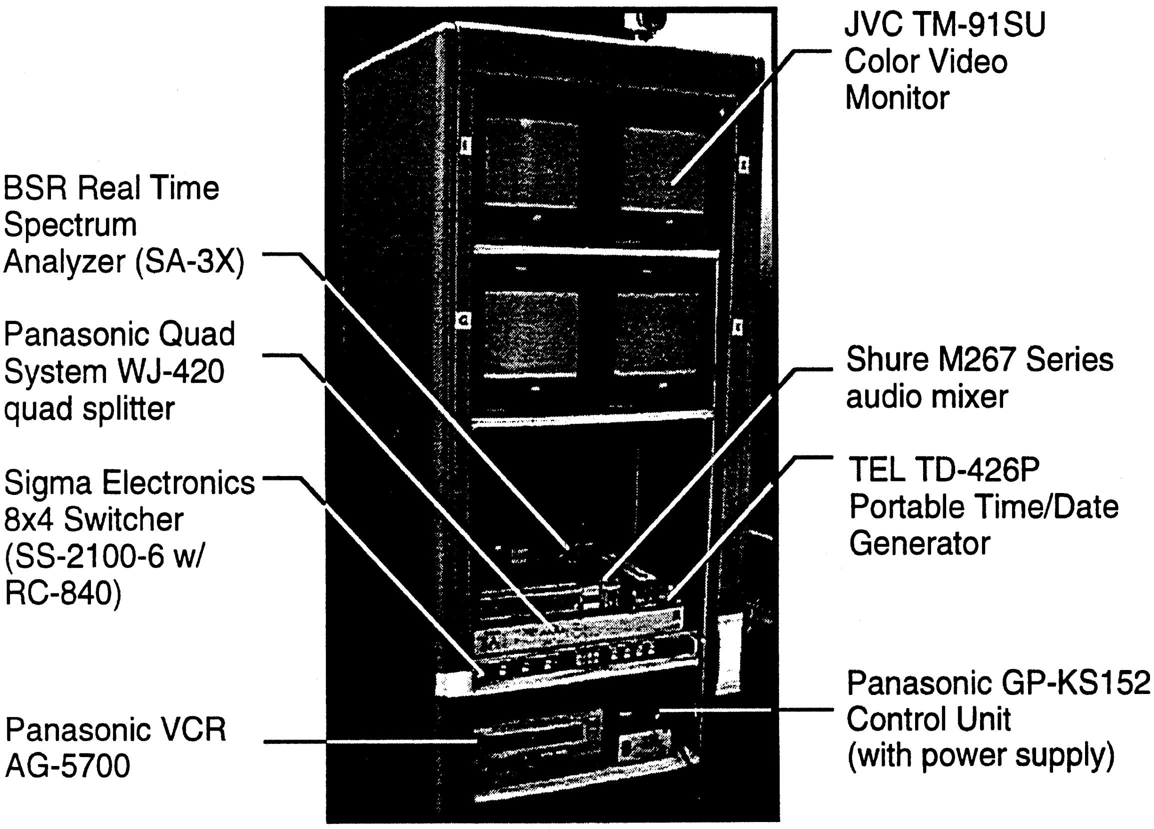


Figure 44. The video rack.

APPENDIX F - PARTICIPANT CONSENT FORM

Subject: ☐ **Date:** ☐

In-Vehicle Menu Study
Participant Consent Form

The primary purpose of this experiment is to investigate how easy it is to control the radio, climate control, and other automotive systems using a computer menu system. Specifically, we are looking at how the location of the control and display, the type of control, and other factors affect the time it takes to use the system. While driving a simulator, you will be asked to turn things on and off by selecting menu items with a knob control or buttons.

A few drivers experience motion discomfort while operating the simulator. Should you feel uncomfortable at any time and for any reason, you may stop the experiment. You will be paid regardless.

The entire study will take approximately two hours to complete. You will be paid $30 for your participation.

We thank you for taking part in this study. If you have any questions at any time, please do not hesitate to ask.

Do you object to being videotaped?

No ☐ **Yes** ☐

I have reviewed and understand the information presented above. My participation in this study is entirely voluntary.

Signature X ______________________________ **Date:** ☐

APPENDIX G - BIOGRAPHICAL FORM

Subject: ☐ **Date:** ☐

In-Vehicle Menu Study
Biographical Form

General Information

Name: ☐

Sex: **Male** ☐ **Female** ☐ **Age:** ☐

Occupation: ☐
(If retired, note your former occupation. If student, note your major.)

Driving Experience

Are you a licensed driver? **Yes** ☐ **No** ☐

How many years have you been driving? ☐

What kind of car do you drive most?
Year: ☐ Make: ☐ Model: ☐

What is your approximate annual mileage: ☐

Simulator Experience

Have you ever driven the UMTRI simulator? **Yes** ☐ **No** ☐

How often do you experience motion sickness?
Never ☐ **Rarely** ☐ **Sometimes** ☐ **Often** ☐

Computer Experience

How many hours per week do you use a computer? ☐

Are you familiar with the Macintosh or Windows operating system?
Yes ☐ **No** ☐

OFFICE USE ONLY — Titmus vision test (Landolt Rings)

Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Location:	T	R	R	L	T	B	L	R	L	B	R	B	T	R
20/:	200	100	70	50	40	35	30	25	22	20	18	17	15	13

APPENDIX H - EXPERIMENTAL DESIGN TABLE

Subj	Sex	Age	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Block 9	Block 10	Block 11	Block 12	Block 13	Block 14
1	M	Y	D C H 1	B C H 2	D K H 3	B K H 4	D C L 5	B C L 6	D K L 3	B K L 5	D C S 1	B C S 6	D K S 2	B K S 4	D C L 7	B K H 8
2	M	Y	D K L 5	B K L 1	D C L 6	B C L 1	D K S 4	B K S 3	D C S 6	B C S 2	D K H 4	B K H 5	D C H 2	B C H 3	D K H 8	B C L 7
3	M	Y	D C S 2	B C S 4	D K S 3	B K S 5	D C H 3	B C H 4	D K H 5	B K H 6	D C L 1	B C L 2	D K L 1	B K L 6	D C H 7	B K L 8
4	M	Y	B K S 1	D K S 5	B C S 3	D C S 4	B K L 2	D K L 6	B C L 3	D C L 2	B K H 1	D K H 6	B C H 5	D C H 4	B K L 7	D C H 8
5	M	Y	B C L 4	D C L 3	B K L 4	D K L 2	B C H 6	D C H 5	B K H 2	D K H 1	B C S 5	D C S 3	B K S 6	D K S 1	B C L 8	D K H 7
6	M	Y	B K H 3	D K H 2	B C H 1	D C H 6	B K S 2	D K S 6	B C S 1	D C S 5	B K L 3	D K L 4	B C L 5	D C L 4	B K H 7	D C L 8
7	M	O	D C H 1	B C H 2	D K H 3	B K H 4	D C L 5	B C L 6	D K L 3	B K L 5	D C S 1	B C S 6	D K S 2	B K S 4	D C L 7	B K H 8
8	M	O	D K L 5	B K L 1	D C L 6	B C L 1	D K S 4	B K S 3	D C S 6	B C S 2	D K H 4	B K H 5	D C H 2	B C H 3	D K H 8	B C L 7
9	M	O	D C S 2	B C S 4	D K S 3	B K S 5	D C H 3	B C H 4	D K H 5	B K H 6	D C L 1	B C L 2	D K L 1	B K L 6	D C H 7	B K L 8
10	M	O	B K S 1	D K S 5	B C S 3	D C S 4	B K L 2	D K L 6	B C L 3	D C L 2	B K H 1	D K H 6	B C H 5	D C H 4	B K L 7	D C H 8
11	M	O	B C L 4	D C L 3	B K L 4	D K L 2	B C H 6	D C H 5	B K H 2	D K H 1	B C S 5	D C S 3	B K S 6	D K S 1	B C L 8	D K H 7
12	M	O	B K H 3	D K H 2	B C H 1	D C H 6	B K S 2	D K S 6	B C S 1	D C S 5	B K L 3	D K L 4	B C L 5	D C L 4	B K H 7	D C L 8
13	F	Y	D C H 1	B C H 2	D K H 3	B K H 4	D C L 5	B C L 6	D K L 3	B K L 5	D C S 1	B C S 6	D K S 2	B K S 4	D C L 7	B K H 8
14	F	Y	D K L 5	B K L 1	D C L 6	B C L 1	D K S 4	B K S 3	D C S 6	B C S 2	D K H 4	B K H 5	D C H 2	B C H 3	D K H 8	B C L 7
15	F	Y	D C S 2	B C S 4	D K S 3	B K S 5	D C H 3	B C H 4	D K H 5	B K H 6	D C L 1	B C L 2	D K L 1	B K L 6	D C H 7	B K L 8
16	F	Y	B K S 1	D K S 5	B C S 3	D C S 4	B K L 2	D K L 6	B C L 3	D C L 2	B K H 1	D K H 6	B C H 5	D C H 4	B K L 7	D C H 8
17	F	Y	B C L 4	D C L 3	B K L 4	D K L 2	B C H 6	D C H 5	B K H 2	D K H 1	B C S 5	D C S 3	B K S 6	D K S 1	B C L 8	D K H 7
18	F	Y	B K H 3	D K H 2	B C H 1	D C H 6	B K S 2	D K S 6	B C S 1	D C S 5	B K L 3	D K L 4	B C L 5	D C L 4	B K H 7	D C L 8
19	F	O	D C H 1	B C H 2	D K H 3	B K H 4	D C L 5	B C L 6	D K L 3	B K L 5	D C S 1	B C S 6	D K S 2	B K S 4	D C L 7	B K H 8
20	F	O	D K L 5	B K L 1	D C L 6	B C L 1	D K S 4	B K S 3	D C S 6	B C S 2	D K H 4	B K H 5	D C H 2	B C H 3	D K H 8	B C L 7
21	F	O	D C S 2	B C S 4	D K S 3	B K S 5	D C H 3	B C H 4	D K H 5	B K H 6	D C L 1	B C L 2	D K L 1	B K L 6	D C H 7	B K L 8
22	F	O	B K S 1	D K S 5	B C S 3	D C S 4	B K L 2	D K L 6	B C L 3	D C L 2	B K H 1	D K H 6	B C H 5	D C H 4	B K L 7	D C H 8
23	F	O	B C L 4	D C L 3	B K L 4	D K L 2	B C H 6	D C H 5	B K H 2	D K H 1	B C S 5	D C S 3	B K S 6	D K S 1	B C L 8	D K H 7
24	F	O	B K H 3	D K H 2	B C H 1	D C H 6	B K S 2	D K S 6	B C S 1	D C S 5	B K L 3	D K L 4	B C L 5	D C L 4	B K H 7	D C L 8

KEY:

Block:
|Menu structure|Control type|Configuration|Stimulus Set|
Note: Blocks 13 and 14 are practice

Menu structure:
D = deep
B = broad

Control type:
C = cursor
K = knob

Configuration:
H = high
L = low
[illegible]

APPENDIX I - ACCEPTABILITY SURVEY

Subject: ☐ **Date:** ☐

In-Vehicle Menu Study
Acceptability Survey

1. What is your overall impression of the computer menu system you just used?

2. Assuming you were going to buy an automobile that has many more features than today's cars but the same amount of space on the dashboard, how would you rank the following alternatives in terms of how comfortable you would be using each one?

   a. A computer menu system similar to the one you used in this study, except that features that are used most frequently (such as volume and temperature) would get separate controls (knobs, buttons, etc.)

   b. A system that understands voice commands (such as "please turn on the air conditioning")

   c. A system that uses knobs, buttons, and switches, but consists of a larger number of them placed closer together than in today's cars

   1st choice ☐ **2nd choice** ☐ **3rd choice** ☐

3. What do you feel are the advantages of using a computer menu system like the one you just used?

4. What do you feel are the disadvantages of using a computer menu system like the one you just used?

APPENDIX J - EXPERIMENTER INSTRUCTIONS

Experimenter Instructions

Before the subject arrives

Label a video tape.
Grab the binder, Bernoulli, Tavern key, video tape, and a pen.

Get a set of forms ready.

Flip the signs.
Plug in the display.
Turn on all the necessary computers.
Turn on the stereo and bass shaker.
Turn on the IP and overhead.
Turn on the camera and remove the lens cap.
Turn on the video cart and insert the video tape.
Load up the simulator.
Press OK on the sound computer.
Open "Driver Interface World."
Load up MenuPlayer.
Set up the display for the 13th block of trials but leave it off.
Set up the control for the 13th block and put the other control in the opposite location.
Turn on the sunroof light.

When the subject arrives

Have the subject fill out the Project Consent Form.
Have the subject fill out the Biographical Form.
Test the subject's vision.
Make sure the subject has all his or her belongings.

Have the subject sit in the car.
Set the timer for two minutes.
Calibrate the simulator.

I'm going to have you drive the simulator for two minutes so you can get used to it. Please try to maintain a speed of 55 miles per hour and stay in the right-hand lane at all times.
Start the simulator.
Start the timer.
Have the subject drive the simulator until the timer goes off.
Pause the simulator.

Tutorial and instructions

As you read earlier, you are going to be using a computer menu system to turn various things on and off or adjust certain settings. Now, let's take a look at the system you will be using.

Turn on the display.

During the experiment, you will be using two different types of menus. The one you see now has four items per menu, the other one (which I will show you in just a second) has eight items per menu.

Switch to the broad menu structure.

To select the menu items, you will be using two different controls throughout the experiment: a knob control and a cursor control. Let's start with the knob control first. To select an item, you need to turn the knob until the black bar moves to the desired item and then press the right-arrow button. Try selecting "Fan/Temp."

Good. You are now in the "Fan/Temp" menu. If you need to go back a menu, press the left arrow. Try that now.

Good. You are now back in the "Main" menu.

Now, I'm going to switch back to the other type of menu.

Switch to the deep menu structure and to the cursor control.

Now, let's try the cursor control. To select an item, push the up and down arrows until the black bar is over the desired item and then press the right-arrow button (just as you did with the knob control). Try selecting "Climate Control."

Good. Now you are in the "Climate Control" menu. If you need to go back a menu, you do the same thing as with the knob control—push the left arrow. Try that now.

Just like last time, you are back in the "Main" menu.

Do you have any questions on what we've done so far?

Running the experiment

Before we begin a practice run, I'm going to explain what to expect. The system will chime and then say a message like, "How would you turn on the radio?" You will then use the control to select menu items until you have selected the requested item. At that point, the system will either beep if you chose the correct item or buzz if you chose the wrong one. Then the system will pause for a few seconds and the process will repeat.

If you don't know where to go to find an item, please take your best guess. Remember, you can always go back a menu if you make a mistake. If you get completely stuck, I can help you, but I will not be able to help you once the real experiment begins. Also, if you don't hear what you are supposed to do, that information will also be displayed at the bottom of the screen.

Do you have any questions?

OK. Before we begin, remember to maintain a speed of 55 miles per hour and stay in the right-hand lane at all times. Please make controlling the car your highest priority, just like you would in real life.

Run Blocks 13 and 14.
Start the tape.
Run Blocks 1 through 12.
Stop and rewind the tape.

Running a block

1. Set up the control and display for the next block (if necessary).
2. Save a simulator data file as *S##B##.sim* and set the time to 1200 seconds.
3. Start the simulator.
4. Run MenuPlayer for Subject ##, Block ##.
5. Pause the simulator.
6. Close the simulator data file.
7. Save the MenuPlayer file as *S##B##.*
8. Set MenuPlayer to the next block.

Before the subject leaves

Have the subject fill out the Acceptability Survey.
Have the subject fill out the appropriate payment form.
Pay the subject.

After the subject leaves

Put the forms in the binder.
Backup the data.
Eject the tape.
Make sure everything is put away and shut down.
Flip the signs.

APPENDIX K - ACTUAL TIMES VERSUS GOMS PREDICTIONS

Table 25. Actual times versus GOMS predictions for the first 32 end nodes by menu structure and control.

		Broad						Deep					
		Cursor			Knob			Cursor			Knob		
#	Menu Item	n	t	t^	n	t	t^	n	t	t^	n	t	t^
1	Volume	4	2.38	4.08	5	2.49	3.86	4	4.20	4.32	6	2.95	4.26
2	Bass	5	5.74	3.95	4	2.88	3.87	3	3.41	4.64	5	4.20	4.67
3	Treble	9	2.85	4.30	9	4.40	4.36	11	4.52	4.75	11	5.57	4.85
4	Balance	9	4.69	4.65	5	4.99	4.45	8	4.04	5.14	10	3.91	5.26
5	Tune	11	5.63	4.91	10	6.20	4.93	10	4.33	4.64	9	5.40	4.67
6	Seek	3	4.43	5.19	2	4.62	5.15	6	6.11	4.96	3	3.97	5.08
7	Preset	8	5.29	5.57	8	6.02	5.56	12	5.54	5.07	10	5.49	5.26
8	Scan	2	4.63	5.50	3	6.31	5.71	5	7.48	5.46	4	6.55	5.67
9	Skip Track	7	2.67	3.95	6	3.02	3.87	12	3.93	4.75	9	4.20	4.85
10	Scan Tracks	9	3.48	3.82	9	3.80	3.88	9	4.12	5.07	9	5.36	5.26
11	Skip Disk	3	3.46	4.17	4	4.50	4.37	2	4.96	5.18	4	6.54	5.44
12	Random Mode	4	5.78	4.52	5	4.51	4.46	5	6.02	5.57	5	5.53	5.85
13	Fast Forward	6	4.23	4.78	4	4.25	4.94	2	4.68	5.14	3	5.67	5.26
14	Rewind	11	4.38	5.06	12	5.08	5.16	6	4.35	5.46	6	5.20	5.67
15	Tape Direction	5	5.64	5.44	4	6.94	5.57	3	4.91	5.57	3	4.62	5.85
16	Eject Tape	5	4.25	5.37	9	4.79	5.72	8	6.85	5.96	9	6.33	6.26
17	Fan Off	9	2.69	4.30	7	3.61	4.36	10	3.71	4.64	8	5.01	4.67
18	Auto	11	3.49	4.17	6	4.04	4.37	9	5.51	4.96	7	4.07	5.08
19	Econ	5	6.37	4.52	5	3.87	4.86	4	4.71	5.07	5	4.86	5.26
20	Fan Speed	6	5.87	4.87	5	5.22	4.95	3	4.25	5.46	6	4.36	5.67
21	Left Temp	4	8.13	5.13	2	6.42	5.43	2	5.25	4.96	4	3.48	5.08
22	Right Temp	10	5.34	5.41	6	6.83	5.65	8	4.64	5.28	10	5.66	5.49
23	Rear Temp	10	5.36	5.79	8	8.77	6.06	6	4.07	5.39	6	5.29	5.67
24	Outside Temp	5	4.56	5.72	1	7.52	6.21	3	7.53	5.78	5	5.24	6.08
25	Panel	2	2.99	4.65	2	4.32	4.45	3	6.82	5.07	5	6.09	5.26
26	Floor	6	3.84	4.52	3	4.44	4.46	4	4.82	5.39	2	6.70	5.67
27	Panel & Floor	11	4.55	4.87	6	5.80	4.95	8	4.95	5.50	6	5.48	5.85
28	Defrost	10	6.05	5.22	5	5.16	5.04	7	4.52	5.89	9	6.24	6.26
29	Fresh Air	6	5.28	5.48	10	5.79	5.52	7	4.73	5.46	6	4.98	5.67
30	Recirculate	4	7.08	5.76	5	5.06	5.74	2	3.92	5.78	2	4.44	6.08
31	Mixed Air	6	5.99	6.14	5	6.39	6.15	5	9.87	5.89	4	5.04	6.26
32	Filtered Air	9	4.94	6.07	7	6.87	6.30	10	7.19	6.28	5	7.18	6.67

n = the number of trials used to calculate the mean for the actual time
t = the actual time in seconds
t^ = the time predicted by the GOMS model in seconds

Table 26. Actual times versus GOMS predictions for the remaining 32 end nodes by menu structure and control.

		Broad						Deep					
		Cursor			Knob			Cursor			Knob		
#	Menu Item	n	t	t^	n	t	t^	n	t	t^	n	t	t^
33	Zoom In/Out	11	5.26	4.91	8	4.73	4.93	9	4.18	4.75	8	3.73	4.85
34	North Up	6	4.52	4.78	4	5.68	4.94	4	8.22	5.07	2	18.57	5.26
35	Heading Up	6	4.21	5.13	4	4.62	5.43	5	8.38	5.18	4	7.85	5.44
36	Show Names	8	4.90	5.48	11	5.95	5.52	6	5.22	5.57	7	5.29	5.85
37	Preset Location	6	4.84	5.74	5	5.61	6.00	2	7.38	5.07	3	5.85	5.26
38	Enter Address	11	7.46	6.02	10	6.65	6.22	8	5.39	5.39	7	5.39	5.67
39	Find Business	5	9.28	6.40	6	4.74	6.63	6	5.72	5.50	5	5.88	5.85
40	Locate on Map	7	6.95	6.33	9	6.35	6.78	5	6.43	5.89	7	8.92	6.26
41	Shortest Route	3	5.97	5.19	5	5.07	5.15	4	4.78	5.18	3	5.96	5.44
42	Fastest Route	10	5.57	5.06	7	6.06	5.16	10	6.12	5.50	12	5.57	5.85
43	No Highways	10	4.31	5.41	9	5.11	5.65	12	5.53	5.61	9	6.99	6.03
44	Scenic Route	3	8.00	5.76	3	5.92	5.74	6	5.16	6.00	5	4.75	6.44
45	Voice Alert	8	5.15	6.02	8	6.55	6.22	8	5.29	5.57	6	9.03	5.85
46	Chime Alert	6	11.35	6.30	4	5.62	6.44	3	5.08	5.89	4	8.26	6.26
47	Voice & Chime	10	7.69	6.68	10	9.14	6.85	10	4.46	6.00	9	4.83	6.44
48	No Alert Sound	4	6.45	6.61	2	10.72	7.00	6	8.30	6.39	5	7.21	6.85
49	Full-Time 4WD	5	9.87	5.57	4	4.07	5.56	3	5.02	5.14	3	4.62	5.26
50	4WD High	11	4.96	5.44	7	4.88	5.57	6	4.77	5.46	6	5.34	5.67
51	4WD Low	8	5.02	5.79	10	6.45	6.06	11	5.68	5.57	8	6.91	5.85
52	2WD	5	4.31	6.14	5	5.34	6.15	1	6.40	5.96	2	4.98	6.26
53	Economy Shift	6	4.62	6.40	5	5.17	6.63	1	5.38	5.46	1	5.72	5.67
54	Normal Shift	9	6.89	6.68	9	7.02	6.85	9	5.73	5.78	6	4.83	6.08
55	Power Shift	6	9.37	7.06	3	5.63	7.26	4	4.97	5.89	2	5.01	6.26
56	Hold Gear	6	6.46	6.99	8	8.75	7.41	8	7.47	6.28	5	5.99	6.67
57	Easy Steering	6	4.60	5.50	5	5.71	5.71	10	4.90	5.57	10	5.57	5.85
58	Light Steering	3	5.44	5.37	3	5.69	5.72	6	5.20	5.89	3	5.58	6.26
59	Med Steering	6	5.06	5.72	2	5.35	6.21	5	7.27	6.00	3	4.93	6.44
60	Firm Steering	5	6.29	6.07	5	5.40	6.30	6	5.14	6.39	4	8.23	6.85
61	Normal Ride	6	5.13	6.33	6	5.97	6.78	5	6.48	5.96	6	4.28	6.26
62	Touring Ride	3	5.24	6.61	3	8.97	7.00	3	7.20	6.28	2	13.82	6.67
63	Sport Ride	6	5.23	6.99	7	5.56	7.41	7	7.11	6.39	10	6.45	6.85
64	Off Road	1	13.65	6.92	2	7.18	7.56	3	5.44	6.78	3	4.27	7.26

n = the number of trials used to calculate the mean for the actual time
t = the actual time in seconds
t^ = the time predicted by the GOMS model in seconds