Technical Report UMTRI-2001-24 June, 2002

Destination Entry while Driving: Speech Recognition versus a Touch-Screen Keyboard

Omer Tsimhoni, Daniel Smith, and Paul Green

umtri
HUMAN FACTORS

Technical Report Documentation Page

1. Report No. UMTRI-2001-24	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Destination Entry while Driving: Speech Recognition versus a Touch-Screen Keyboard		5. Report Date June, 2002 6. Performing Organization Code account 376015
7. Author(s) Omer Tsimhoni, Daniel Smith, and Paul Green		8. Performing Organization Report No. UMTRI-2001-24
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.
12. Sponsoring Agency Name and Address Mitsubishi Motors Corporation 1 Nakashinkiri, Hashime-Cho, Okazaki, Aichi, 444-8501, Japan		13. Type of Report and Period Covered 8/00 – 7/01 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

To determine the effect of several destination-entry methods on driving performance, as a function of driving workload, 24 participants drove a simulator on roads with curves of different radii while entering street addresses into a simulated in-vehicle navigation system. Three destination-entry methods were examined: 1) speech recognition by word, 2) speech recognition by character, and 3) typing on a touch-screen keyboard. For each method, driving performance, glance behavior, task partitioning, and subjective evaluation were examined.

While driving, speech recognition by word yielded the shortest task completion time (15.3 s), followed by speech recognition by character (41.0 s), and then the touch-screen keyboard (86.0 s). Overall, driving performance degraded when entering destinations, especially with the touch-screen keyboard, and when driving workload increased. The participants made 21% shorter glances to the navigation display, and 40% longer glances at the road scene as a function of increasing driving workload. Similarly, touch-screen keyboard-entry behavior was significantly affected by an increase in driving workload and age. When driving, pauses between fields of an address increased 57% and pauses between groups of characters increased 27%. These results confirm the risks associated with destination entry using a touch screen and suggest the use of speech recognition. However, speech recognition systems with less than perfect accuracy and visual feedback are not risk free.

17. Key Words		18. Distribution Statement	
Driving, Visual Demand, Destination Entry, Workload, Touch Screen, Speech Recognition, ITS, Human Factors, Ergonomics, Safety, Usability, Telematics		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 104	22. Price

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

DESTINATION ENTRY WHILE DRIVING: SPEECH RECOGNITION VERSUS A TOUCH-SCREEN KEYBOARD

UMTRI Technical Report 2001-24
Omer Tsimhoni, Daniel Smith, and Paul Green

University of Michigan Transportation Research Institute
Ann Arbor, Michigan, USA

1 Issues

1. Effect of destination entry method (speech recognition by word, by character, and touch-screen typing) on:
   - task completion time (and errors)
   - concurrent driving performance
   - perceived difficulty and perceived safety
2. Effect of driving workload (while parked, on a straight road, a moderate curve, and a sharp curve) on:
   - destination-entry task performance
   - glance behavior
   - task partitioning (character entry)

2 Test Method

Destination Entry

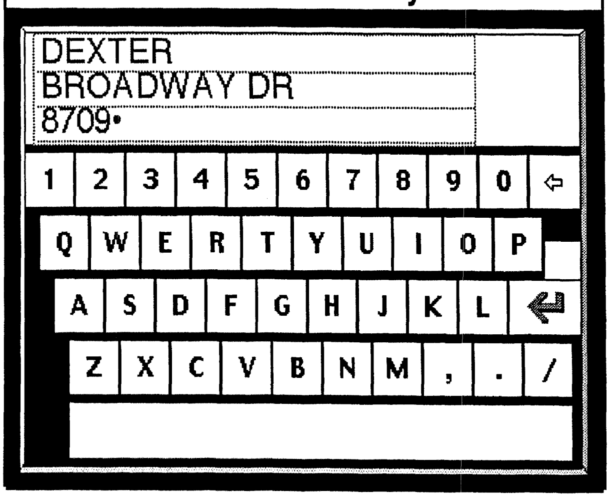
Keyboard	Typed on touch-screen
Speech Recognition	
Word dictation	" DEXTER "
Character spelling	"D" "E" "X" "T" "E" "R"

Keyboard Entry Method

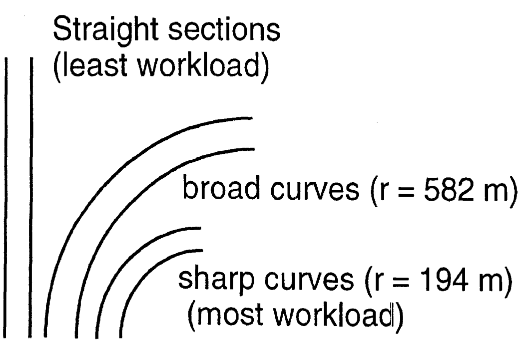
24 Subjects

	Female	Male
Young (20-29)	6	6
Old (65-72)	6	6

20 Character Address entered on modified **QWERTY keyboard**



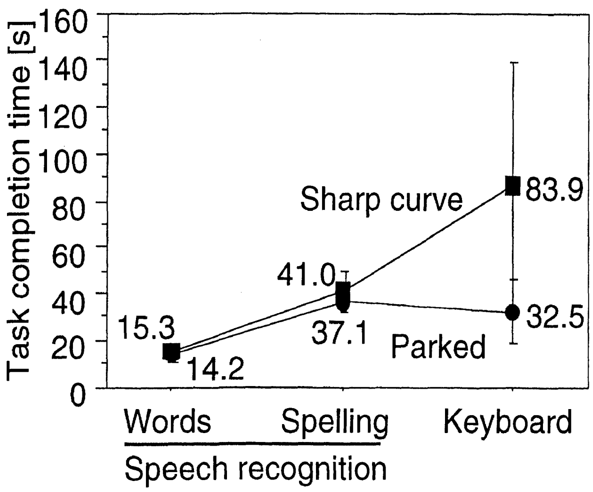
3 Driving Workload Levels



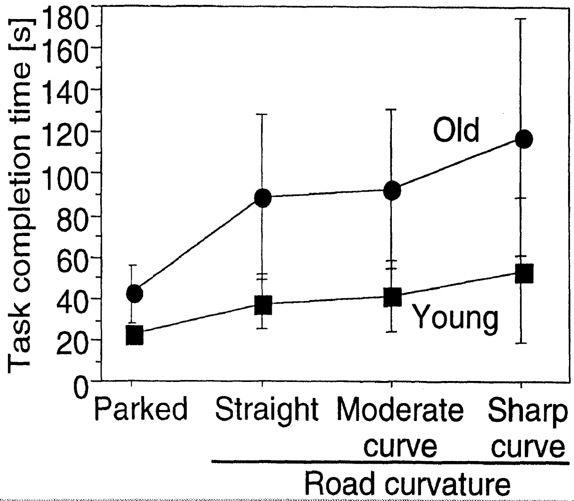
3 Results

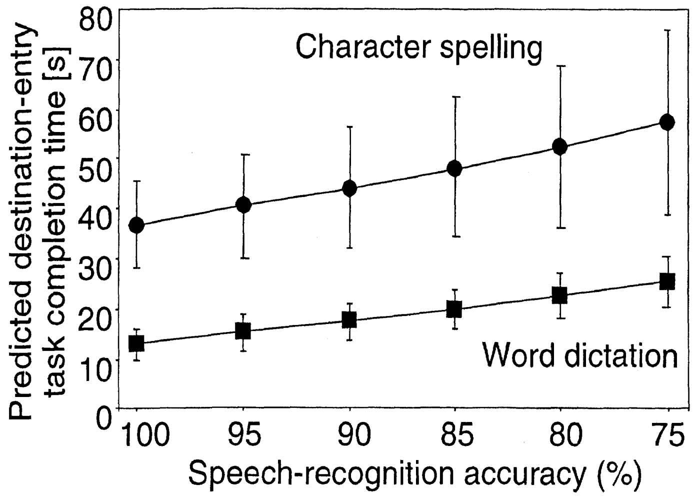
Destination-Entry Performance

♦ Task completion time while driving was shortest in word mode, and longest in keyboard-entry.



♦ Task completion time with the touch-screen keyboard increased with increasing driving workload.



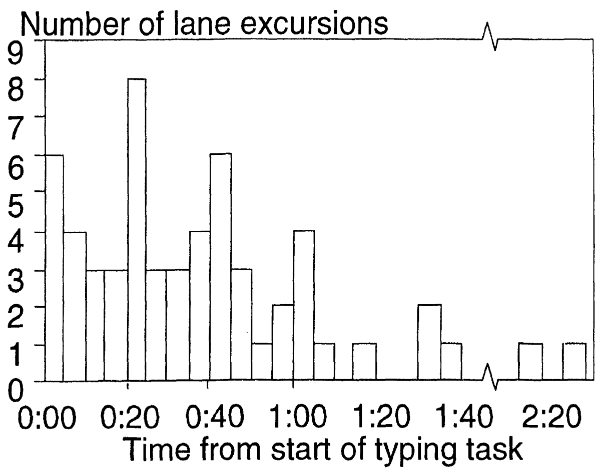


Mean entry-time per key

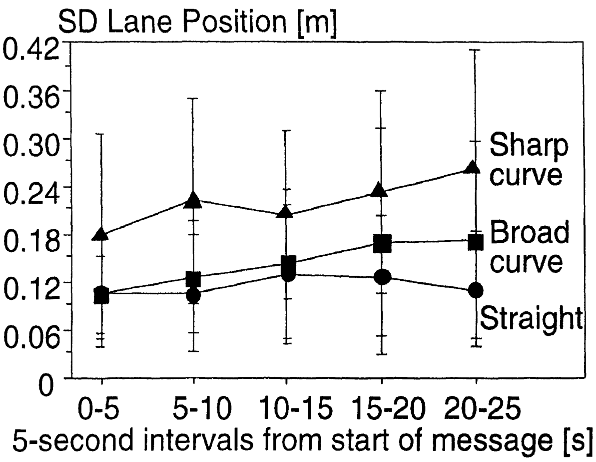
	Enter Key	Space Key	Letters/ Numbers
Parked	1.3	1.1	1.1
Straight	2.6	2.0	1.7
Moderate Curve	2.6	2.2	2.1
Sharp Curve	2.4	2.7	2.3

Driving Performance

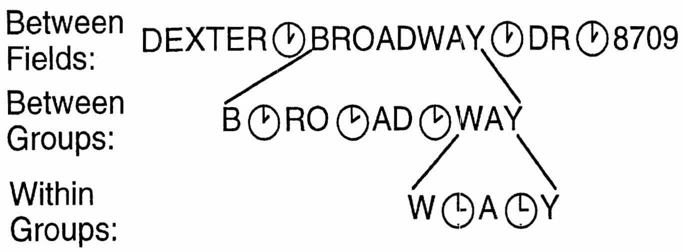
♦ Most first lane departures occurred within the first minute of the entry task.



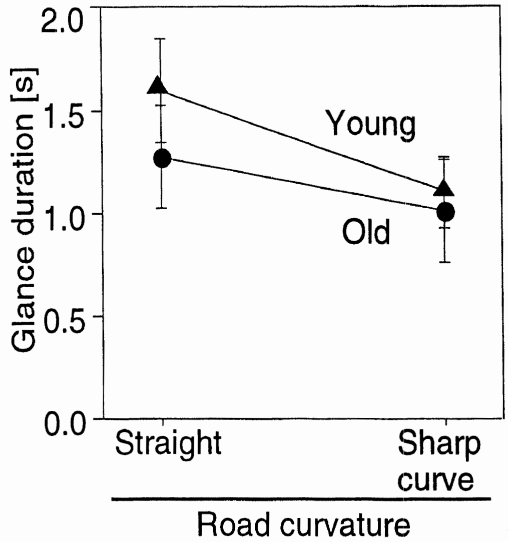
♦ Standard deviation of lane position increased with elapsed time.



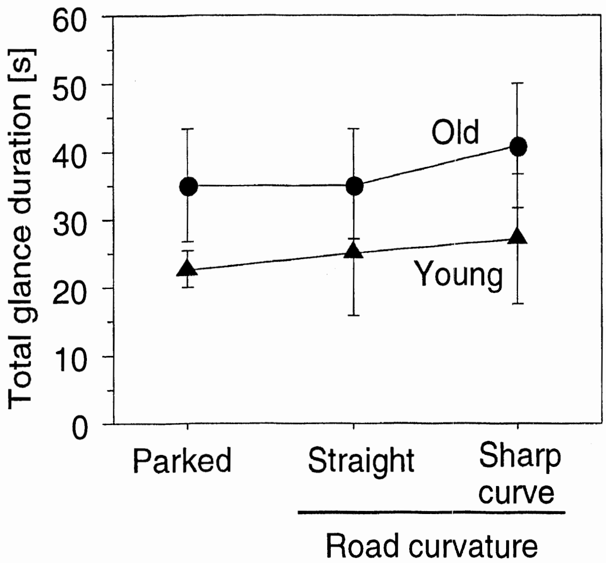
Task Partitioning



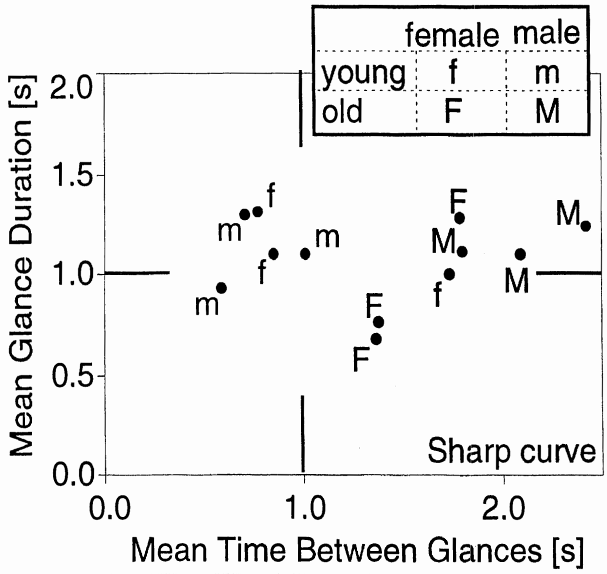
♦ Pauses between fields (+57%) and between groups (+27%) increased as road curvature increased.



♦ Total glance duration at the display remained relatively constant under all workload levels.

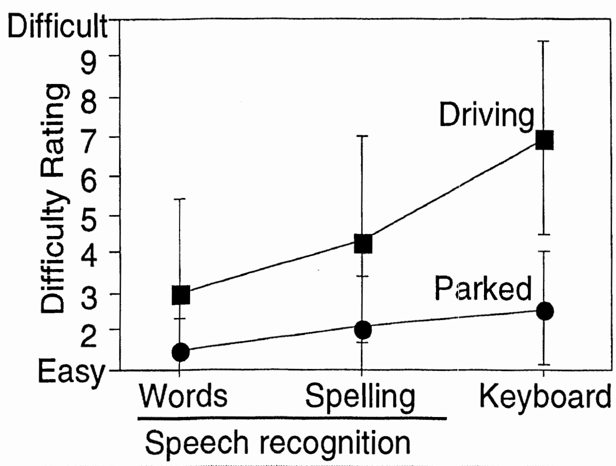


♦ Glances at the display were consistent for all participants. Glances at the road by young participants were shorter.



Post-Test Subjective Rating

♦ Driving rated more difficult than parked, and keyboard entry rated most difficult.



♦ Keyboard entry rated least safe, and no method rated extremely safe.

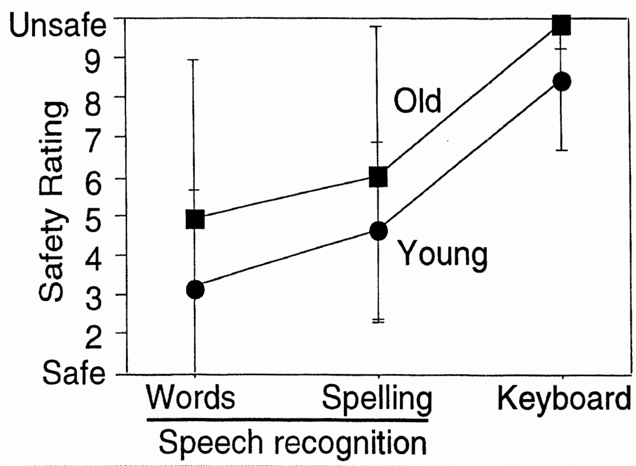


TABLE OF CONTENTS

INTRODUCTION

In recent years there has been considerable interest in providing telematics in motor vehicles. In the near future, vehicle manufacturers expect that a significant share of their profits will be associated with the sales of telematics features and services (Richardson and Green, 2000). In fact, a few believe that manufacturers will make no profits on sales of the basic vehicle, only on telematics equipment and services.

For telematics to see broad market penetration they must be useful, easy to use, and safe to use. The goal of usefulness has been achieved for navigation, at least in Japan, where sales of navigation systems has been substantial.

To aid the development of easy-to-use systems, several sets of usability and safety guidelines have been developed, such as Green, Levison, Paelke, and Serafin (1995), Ross, Vaughan, Engert, Peters, Burnett, and May (1995), British Standards Institution (1996), Campbell, Carney, and Kantowitz (1998), Japan Automobile Manufacturers Association (2000), and Society of Automotive Engineers (2000). (For a summary, see Green, 2000, 2001b and http://www.umich.edu/~driving/guidelines.html, the UMTRI Driver Interface web site.) Those guidelines have a body of research as their foundation (e.g., Wierwille, Antin, Dingus, and Hulse, 1988; Parkes and Franzen, 1993; Noy, 1997; Wakita, and Terashima, 1999). A major shortcoming of that research has been insufficient consideration of the effects of workload and the failure to quantify workload when it is examined.

Workload is of concern because of the crash risk associated with overload. If, for example, a driver is spending a substantial amount of time looking at and manipulating an in-vehicle interface, then the time available to look at the road is significantly reduced, which increases crash risk. Similarly, if drivers are engaged in a phone conversation and their attention is focused on that conversation, then their awareness of the driving situation is reduced, again increasing crash risk. For these reasons, the topics of driver distraction and driver workload are of considerable interest.

Workload Assessment

Tsimhoni and Green (1999) utilized the visual occlusion method to assess the visual demand of driving on curves. They found a high correlation between the fraction of time drivers needed to see the road and the inverse radius of curvature. In addition, they also determined that the visual demand for a curve began to increase from the steady state value prior to a curve at a distance of about 150 meters, peaked at or near the curve entry, and leveled within about 100 meters of the peak. Quantifying the visual demand of driving these curves allowed for controlling the difficulty of driving in the current experiment.

On a follow-up study, Tsimhoni, Yoo, and Green (1999) conducted an experiment to explore how map-reading and driving performance were simultaneously affected by the visual demand of driving. As visual demand increased, driving performance declined. Further, participants made more and shorter glances to the display, but waited longer between glances. The net effect was that eyes-off-the-road time decreased slightly as visual demand increased.

A logical complement to the examination of display-intensive tasks and the role of workload in the previous experiment was to extend the coverage to control-intensive tasks (Green, 1999a,b). For telematics, destination entry is of particular concern and was therefore the focus of the current experiment.

Display-Intensive Tasks

Destination entry has been the topic of several prior UMTRI studies (Steinfeld, Manes, Green, and Hunter, 1996; Manes, Green, and Hunter, 1998; Nowakowski, Utsui, and Green, 2000; Society of Automotive Engineers, 2001). Steinfeld et al. examined the time to enter and retrieve destinations from a simulated Siemens Ali-Scout navigation system in a vehicle mockup. Median entry times were 51.5 s, with a twofold increase from younger (ages 20-30) to older (over 65) participants. Manes et al. carried out further analysis of the entry and retrieval tasks. They determined the mean times for each type of keystroke and developed a four-step method for estimating task completion times based on the number and type of keystrokes. Nowakowski et al. studied the usability of a laptop navigation system with a keyboard and a remote control and how driving affected task completion times. They found that driving slightly increased task completion times. A high correlation between task completion times while parked and while driving was found. They compared measured keystroke times with estimations from SAE J2365 (Society of Automotive Engineers, 2001) and found reasonable agreement.

One result of these studies on destination entry has been the development of SAE Recommended Practice J2365, a method for estimating navigation system task times (Society of Automotive Engineers, 2001). Times for elemental actions were derived from the keystroke-level model described by Card, Moran, and Newell (1980, 1983), data from Olson and Nilsen (1997-1998), data from Manes, Green and Hunter (1998), and reach times from Methods Time Measurement-1 (MTM-1)(Schwab, 1971). The standard provides time estimations for basic elements such as reaching to a display hitting a cursor, and typing in characters or numbers. The standard provides a practical way for estimating task completion times in an automotive context.

For a more detailed summary of prior UMTRI research on workload assessment, display intensive tasks, and manual input tasks, see Appendix A.

Speech Recognition

Although manual data entry while driving has been studied extensively, it is not the only means available for entering data into in-vehicle systems. In recent years, speech recognition systems have become available in some vehicles. Since voice interaction with in-vehicle devices is thought to interfere less with the driving task than manual entry, voice input is an excellent candidate for data-entry tasks. Optimally a speech recognition system should not require any visual feedback. However, some systems would still provide visual information to the driver. This is most likely to occur in after-market systems that already have some form of visual feedback (such as a menu or address display) and are fitted with a speech-recognition system to act as a replacement for the keyboard.

There is considerable data in the literature examining voice interaction with mobile telephones while driving (see Goodman, Bents, Tijerina, Wierwille, Lerner, and Benel, 1997 for a review of all mobile phone use while driving), but virtually none regarding data-entry tasks or the comparison between manual and voice data entry. The most closely related experiment (Malkin and Christ, 1985) involved a comparison of voice and keyboard data entry for a helicopter navigation task. Following the experiment, each of the 12 subjects were asked to state their system preference for each of 3 flight conditions, and 83.3% of responses indicated that speech was slightly or much better than manual data entry. Overall, the authors found that while keyboard data entry was fastest ($p<0.0027$), voice data entry was perceived as requiring less effort and it reduced overall pilot workload.

In the current experiment, a comparison is made between manual and voice data entry as a means for destination entry while parked and while driving. To maintain compatibility with some existing systems, and to allow for comparison of the entry task independent of the feedback process, all methods showed the address being entered on the navigation display.

Since all current speech recognition systems are less than 100% accurate, the consequences of recognition errors were of interest. A controlled accuracy rate of 92% per word was used in this experiment, simulating the performance of contemporary systems, which are expected to be 90%-95% accurate (Dr. T. Kuhn, Temic Teleunken, personal communication, March, 2000). Using the error correction times, an attempt was made to predict the effects of other values of error rates on performance.

Thus, to assist in further development of design and performance standards for telematics, the following issues were examined.

Issues This Study Addresses

1. How does the total time for drivers to enter destinations using manual and voice methods compare as a function of the driving workload? (Which is best and when?)

   Total entry time will be shorter using voice recognition in word mode, followed by spelling mode and keyboard.

2. When using a touch-screen keyboard to enter addresses, how do the inter-item intervals (mean time and variability) vary as a function of the driving workload? How does the pattern of pauses differ from the single task (non-driving) situation?

   The times between field times (e.g., between the city name and the street name) were expected to increase with workload. Within fields, the pauses between character chunks, but not the times between successive characters within groups, should increase.

3. What is the pattern of glances for destination entry using a touch-screen keyboard and how does it change as a function of driving workload?

   We expect glances away from the road on sharp curves to be shorter and the intervals of glances back at the road to be longer than on straight, less demanding roads.

4. How much does the addition of the manual and voice input tasks degrade driving performance for each level of workload?

   Both will degrade driving, but the degree of degradation will be greater for manual entry. The threshold for manual entry will be reached at lower levels of workload, leading to an interaction between entry method and workload.

5. How did participants rate the difficulty and safety of the tested entry methods?

6. How does entry performance and behavior vary as a function of driver age and gender?

   Consistent with previous studies, entry times for older drivers will be double those of younger drivers.

Measures of interest for the above questions include:

- Time between characters, time between words, task completion time
- Driving performance (standard deviation of lateral position, excursion rate, standard deviation of steering wheel angle)
- Ratings of difficulty by the participants

TEST PLAN

Overview

Participants drove a simulator on roads with curves of several radii while entering addresses on a navigation system using one of three methods: (1) touch-screen keyboard typing, (2) speech recognition – character spelling or (3) speech recognition – word dictation. Speed and accuracy of address entry, detailed measures of driving performance, and subjective ratings of difficulty and safety were recorded and analyzed.

Test Participants

Twenty-four licensed drivers participated in this experiment, 12 younger (ages 20 to 29, mean of 24) and 12 older (ages 65 to 72, mean of 69). Each age bracket included 6 men and 6 women. Participants were recruited from the UMTRI participant database and were paid $40.

Participants reported driving 3,000 to 25,000 miles per year; on average, they drove slightly more than the mean for U.S. drivers (11,000 miles per year). No participant had a professional driving license. Five younger and 3 older participants reported being in an accident within a 5-year period prior to the study. Table 1 summarizes characteristics of the participants.

Participants' vision was tested using a vision tester (Optec 2000, Stereo Optical Inc.) for far visual acuity, near visual acuity, and color vision. All had far visual acuity of 20/40 or better, as required by Michigan state law for driving (day and night). All had near visual acuity of 20/70 or better. No color deficiencies were detected.

Only 2 of the participants (both young men) had used an in-vehicle navigation system before, but all understood what a navigation system was.

Table 1. Participant information

		Young		Old		Young	Old
		Women	Men	Women	Men		
	Mean age (years)	25 ± 3	23 ± 3	69 ± 2	70 ± 2	24 ± 3	69 ± 2
Driving	Mean annual mileage	10,000	17,000	11,000	11,500	13,500	11,250
	Participants with ≥1 'accidents' in last 5 yrs	2	3	2	1	5	3
Vision	Far visual acuity range	20/13-20/22	20/13-20/22	20/17-20/40	20/15-20/40	20/13-20/22	20/15-20/40
	Near visual acuity range	20/13-20/18	20/13-20/30	20/35-20/50	20/18-20/70	20/13-20/30	20/18-20/70
Typing	Keyboard typing speed (words/min)	48 ± 15	37 ± 20	33 ± 18	17 ± 9	42 ± 18	25 ± 16
	Touch-screen typing speed (words/min)	14 ± 2	11 ± 4	8 ± 2	8 ± 2	13 ± 3	8 ± 2
	Mean hours a day typing	1.2	2.1	0.8	0.3	1.7	0.6
	Median age learned typing (years)	13	15	16	20	14	18

Typing skills

Because this experiment involved manual data entry, typing skill was examined as a potential discriminating characteristic. All participants could type, but their typing skills varied. Typing skill level was not controlled in recruiting participants because older men who could touch type were unavailable. Most participants reported typing for half an hour or more daily. Six of the participants did not type on a regular basis (3 older men, 2 older women, and 1 young man).

Typing speed and accuracy of a sample text were measured at the beginning of the experiment in a 1-minute typing test. Participants typed an excerpt from a story (Appendix I) using an Apple extended keyboard II connected to a Macintosh 7100/80 computer. The experimenter noted the typing method. All older men used the hunt-and-peck typing method. In each of the other 3 age/gender subgroups, 2 participants used the hunt-and-peck method, and 4 were touch typists. Typing speed ranged from 8 to 69 correct words/minute. In an additional typing test using the touch-screen keyboard (Elotouch CTR 2310MX) installed in a vehicle mockup, typing speed ranged from 6 to 17 correct words per minute, about a quarter of the normal typing speed.

Equation 1 and Figure 1 show the relationship between touch-screen and keyboard typing speed.

Touch-screen typing speed [words/min] = (1)
6.56 + 0.11 * keyboard typing speed [words/min]; $R^2 = 0.34$

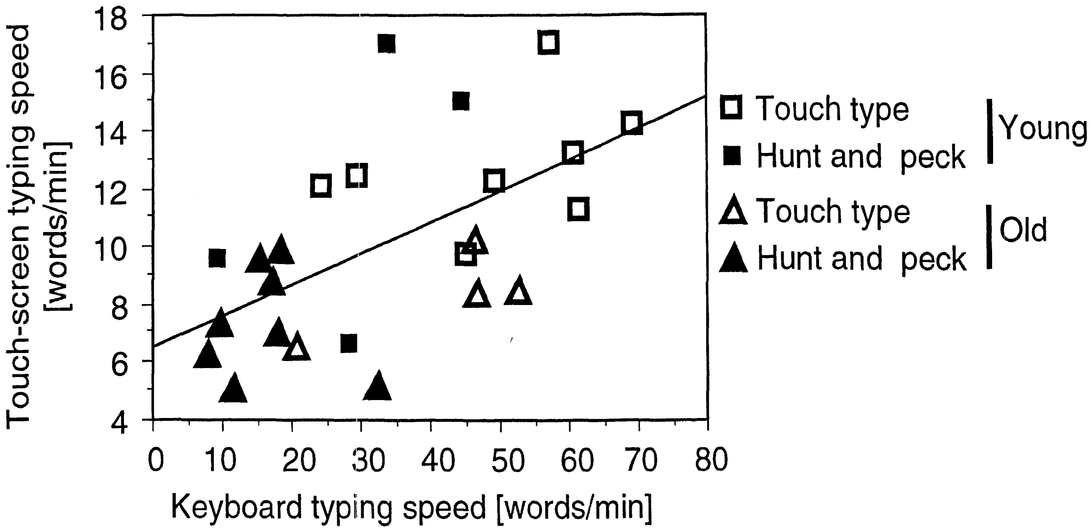


Figure 1. Touch-screen and keyboard typing speed by age and typing method (also see Equations 1, 2, and 3)

To further test the relationship between typing speed on a keyboard and on the touch screen, a stepwise regression of touch-screen typing speed versus keyboard typing speed, age, and typing method was peformed. Age ($p<0.05$) was the only independent variable included in the final regression (Equation 2). Keyboard typing speed and typing method were not significant and therefore were excluded from the regression. A multiple regression that forced the 3 independent variables into the equation resulted in Equation 3. To put these equations in perspective, an average non-typist's speed is 40 words/min (Card, Moran, and Newell, 1983).

Touch-screen typing speed [words/min] = 15.17 - 0.11 * age [years]; $R^2 = 0.49$ (2)

Touch-screen typing speed [words/min)]= 11.23 - 0.084 * age [years] + (3)
0.07 * keyboard typing speed [words/min]+ 0.85 * Hunt and peck; R^2=0.51
[hunt and peck =1, touch = 0]

Based on Equation 3, typing speed on a touch screen was slower for older participants (decreasing 0.08 words/min per year from 12.6 to 7.7 words/min), slightly faster for participants whose keyboard typing was faster (0.7 words/min on the touch screen for every 10 words/min on the keyboard), and it was faster for those who used the hunt-and-peck method (by 0.85 words/min) than those who touch typed.

Experimental Design

This study examined 9 of the 12 combinations of 3 methods (touch-screen keyboard typing, speech recognition – character spelling, speech recognition – word dictation) with 4 levels of driving workload (parked, straight, moderate curve, sharp curve). To keep the duration of the experiment reasonable, 3 conditions out of the 12 possible combinations were not examined (Table 2). Since driving workload did not appear to affect performance in a few pilot runs while in speech recognition mode, the medium levels of driving workload for speech recognition were dropped.

Table 2. Experimental design

		Driving workload (curvature in degrees)			
		Parked	Straight road (0)	Moderate curve (3)	Sharp curve (9)
Entry method	Speech recognition – words (dictation)	✓			✓
	Speech recognition – characters (spelling)	✓	✓		✓
	Touch-screen keyboard	✓	✓	✓	✓

Participants in each age/gender subgroup were randomly assigned to one of two groups. The first group performed the keyboard task before the speech recognition task, while the second group performed the keyboard task after the speech recognition task. Within each of these groups, the order of curvature was manipulated following a Latin square design so that each of the 3 road curvatures appeared first, second, or third for exactly 1 participant in each age/gender subgroup.

Table 3 lists the dependent and independent measures analyzed in this experiment.

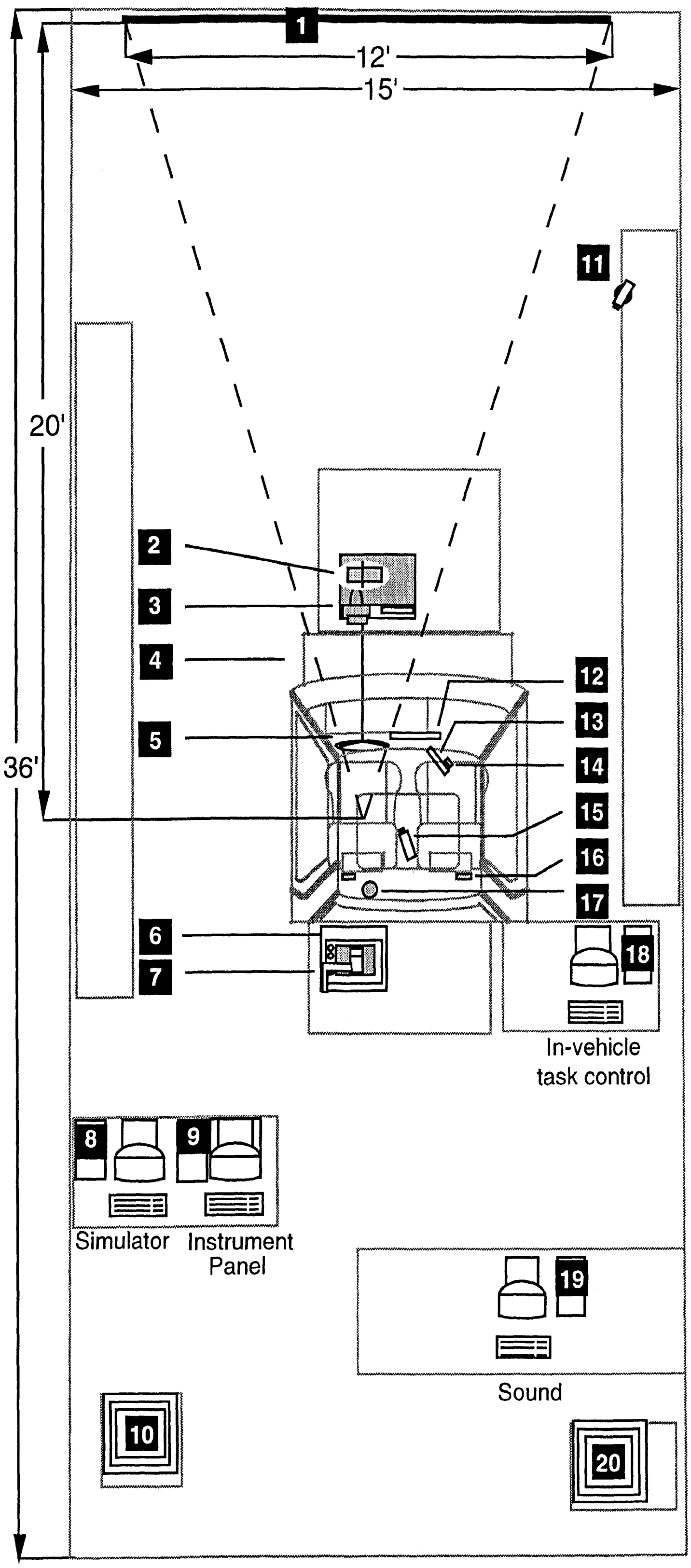
Table 3. Dependent and independent measures

Independent measure (number of levels)		Dependent measure
Entry method (3) Driving workload (4) Age (2) Gender (2) (Participant nested within Age and Gender)	Destination entry	Task completion time Inter-item timing (mean, SD) Number of errors by type
	Driving performance (Lateral control) (Longitudinal control)	Lane departures (count) Lateral position (mean, SD) Yaw angle (mean, SD) Steering angle (SD)
		Headway (mean, SD) Forward velocity (mean, SD) Accelerator position (SD)
	Glance behavior	Total glance duration (mean, SD) Glance duration (mean, SD) Time between glances (mean, SD) Number of glances (mean, SD)
	Task partitioning (Character entry)	Time between fields (mean, SD) Time between groups (mean, SD) Time within groups (mean, SD)
	Subjective ratings	Rating of difficulty (modified Cooper-Harper scale)

Test Materials and Equipment

Driving simulator

The experiment was conducted in the UMTRI Driver Interface Research Simulator, a low-cost driving simulator based on a network of Macintosh computers (Olson and Green, 1997). The simulator (Figure 2) consists 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, drivetrain, tire, and wind noise), a sub-bass sound system (to provide vertical vibration), a computer system to project images of an instrument panel, and other hardware. The projection screen, offering a horizontal field of view of 33 degrees and a vertical field of view of 23 degrees, was 6 m (20 ft) in front of the driver, effectively at optical infinity. The simulator collected driving data at 30 samples per second. For this experiment, road segment number, time, lateral position, steering wheel angle, forward velocity, gas pedal position, and brake pedal position were analyzed.



1 8'X10' projection screen with 3M hi-white encapsulated reflective sheeting

2 PMI Motion Technologies ServoDisk DC motor (model 00-01602-002 type U16M4) with Copley Controls Corp. controller (model 413) and power supply (model 645)

3 Sharp color LCD projection system (model XG-E850U) for instrument panel

4 1985 Chrysler Laser mockup with simulated hood

5 3-spoke steering wheel

6 3M overhead projector (model 9550)

7 Sharp computer projection panel (model QA-1650)

8 Power Macintosh G3

9 Macintosh 7100/80AV

10 Video rack

11 Panasonic low level light camera (model WV-BP510) with TV zoom lens (WV LZ 81)

12 Touch-screen (Elotouch 1225L)

13 5" TFT display (AM-064P)

14 Panasonic .5" micro-camera (GP-KS162)

15 Panasonic low level light camera (model WV-BP550)

16 Sony speaker system (SR6-48)

17 Bass shaker

18 Power Macintosh 9500/120

19 Macintosh Quadra 840

20 Audio rack

Figure 2. Plan view of UMTRI's Driver Interface Research Simulator

Simulated roads

The simulated roads were designed to impose 3 levels of workload by manipulation of road curvature (straight sections, moderate curves [3 degrees of curvature, 582 m radius], and sharp curves [9 degrees of curvature, 194 m radius]). These curve radii were chosen based on results from a previous study (Tsimhoni and Green, 1999), in which the visual demand of these curves was quantified. A linear increase in visual demand was found for curves of 3, 6, 9, and 12 degrees of curvature. In addition, the visual demand at the beginning of curves was higher and decreased to a steady state after approximately 150 m. Therefore, to avoid drastic changes in visual demand within conditions in the current study, curves were designed to be long enough to maintain constant visual demand values. Furthermore, the address entry tasks were presented no earlier than 200 m after the beginning of curves.

Both lanes of the two-lane road were 3.66 m (12 feet) wide. Traffic consisted of five vehicles: the participant's vehicle, a lead vehicle driving in the right lane, and 3 additional vehicles driving in the left lane (Figure 3). The participant was instructed to drive on the right lane at a comfortable distance behind the lead vehicle, which maintained a constant speed of 72.5 km/hr (45 mi/hr). The left-lane lead vehicle drove next to the lead vehicle at a variable speed from 71 to 77 km/h (44 to 48 mi/hr). The second left-lane vehicle drove exactly 4 seconds behind it. The trailing vehicle in the left-lane was a police car that trailed 8 seconds behind the lead vehicle. A typical view of the road, as seen by the participant, is shown in Figure 4.

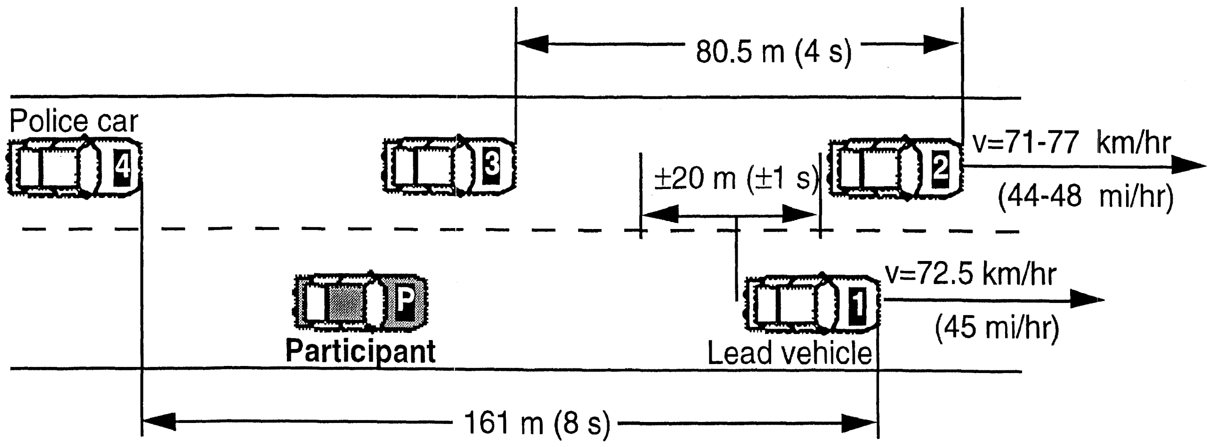


Figure 3. Typical layout of traffic

Figure 4. Typical simulator road scene

Memo display

Addresses appeared on a memo display (5-inch diagonal TFT AM-064P, Advanced Video and Communication Inc.) located to the right of the driver at arm's length. The memo display simulated a low-cost address book or a piece of paper on which an address may appear. Characters on the display were relatively large (0.017 radians, 57 arcmin) to allow easy reading. The center of the memo display was 29±2 degrees below the horizontal line of sight and 56±2 degrees to the right of the center (Figures 5 and 6). This position was chosen to allow the experimenter to distinguish between looks to the memo display and looks to the touch-screen keyboard.

Touch-screen keyboard

A 13-inch touch-screen display (Elotouch 1225L), located in the center console of the vehicle, was used to simulate an entry module of a navigation system. To reduce the active area to a size more likely for a production vehicle, a rectangular 7-inch diagonal opening (4:3 horizontal to vertical ratio) was cut into a black cardboard cover. The center of the touch screen was 23±2 degrees below the horizontal line of sight and 30±2 degrees to the right of the center (Figures 5, 6, and 7).

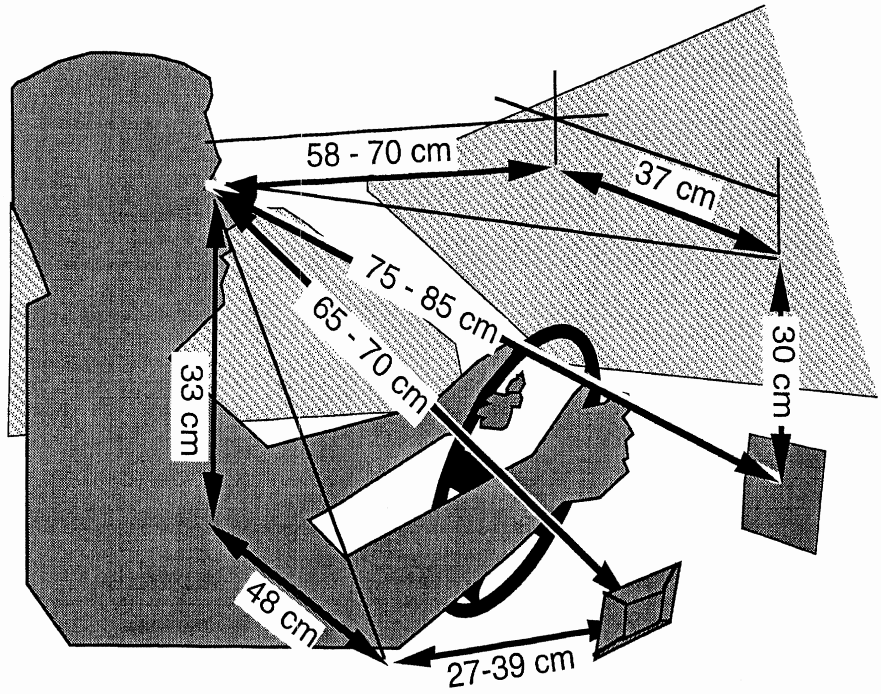


Figure 5. The location of the in-vehicle displays

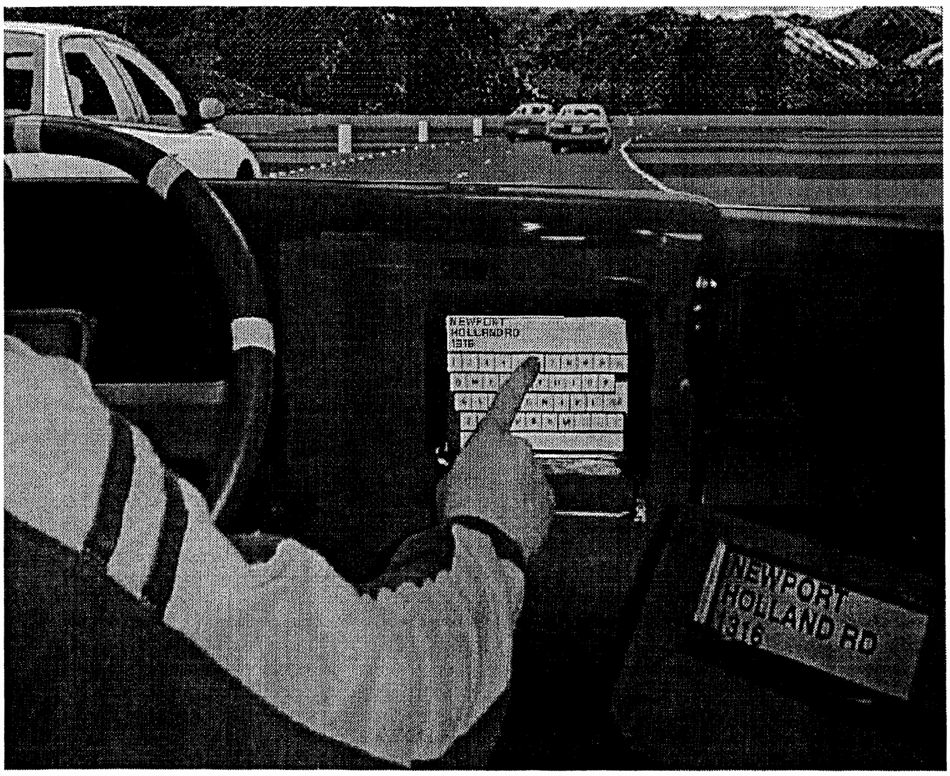


Figure 6. Driver's view of the road and touch screen
(Note: The touch-screen keyboard and road scene were enhanced in this image for clarity)

The touch-screen keyboard was implemented as a standard QWERTY based on a study of reduced-size touch-screen keyboards (Coleman, Loring, and Wiklund, 1991). Coleman and her colleagues compared three types of keyboards (a standard QWERTY, a matrix QWERTY, and alphabetical matrix) on a 5-inch diagonal touch-screen. They found that speed and error rates, as well as subjective preferences, all supported the use of the standard QWERTY keyboard. The keyboard in the current study followed their design but was larger, with a 7-inch diagonal touch screen. Keys were 12.7 mm high and 12 mm wide and were spaced 0.7 mm apart.

Table 4 shows the sizes and visual angles of the keys, the characters on the keys, and the characters on the address display.

Figure 7. Touch-screen keyboard

Table 4. Visual angles of the in-vehicle displays

	Display	Distance from eye [mm]	Height			Sample
			mm	Radian	arcmin	
Touch screen	Address display	800±50	6.3	0.008	27	DEXTER•
	Rectangular key		12.7	0.016	55	
	Characters in keys		4.8	0.006	21	Q W E R
Memo	Address display	670±50	11.1	0.017	57	**DEXTER**

Speech recognition

The "Wizard of Oz" method (e.g., Green and Wei-Haas, 1985) was used to simulate a high-accuracy speech-recognition system. The experimenter, acting as the speech-recognition system, used keyboard shortcuts to present words on the navigation display in response to words or letters the participant said. This method was chosen because: (1) a high accuracy, word level, hardware-based speech recognition system was not readily available and would have been very costly, (2) the accuracy of commercial recognition systems for addresses, and especially in spelling mode, is well below the expected accuracy for future in-vehicle systems, (3) the Wizard of Oz method allowed full control of the accuracy of recognition, and (4) participants did not require a training session to use the speech-recognition system.

When the recognition system was in character spelling mode, the participant was instructed to spell out each of the 4 elements of the address (city, street name, street suffix, and number). After each character was spoken, the experimenter pressed the space bar, which presented a keyboard click sound to the participant (after about 300 ms), confirming the character was heard. After the whole item was spelled out, the experimenter used a shortcut (e.g., Control-1, Control-2) to display it on the navigation system. In each block of 6 addresses, 2 addresses were predetermined to include an unrecognized word. The participant was instructed to delete the unrecognized word as soon as it was detected by saying "scratch that" and repeating the correct word. The "scratch that" phrase is used by a popular speech recognition system for correcting errors in English (IBM Via Voice, version 2.03).

After the full address was spelled out correctly, the system played a chime used in automotive speech interface prototypes and accepted the address. If there was a spelling error in the address, the system played an "oops" sound and the participant had to repeat the entry using the "scratch that" method described above.

The accuracy of the speech recognition system was approximately 92%, simulating the performance of contemporary systems, which are expected to be 90%-95% accurate (Dr. T. Kuhn, Temic Teleunken, personal communication, March, 2000). In spelling mode, the expected accuracy for one character is as low as 40% depending on the character because without the context, it is very difficult to discriminate between characters that sound similar. However, a 90% - 95% recognition rate can be reached for a set of characters based on the context of the word or on an existing dictionary.

Each block included 24 items (6 addresses with 4 items each: city, street name, street suffix, and street number), of which 2 items were unrecognized. The addresses that were unrecognized were the same for all the participants. They were randomly chosen prior to the experiment and scripted into the shortcuts that the experimenter used.

Several measures were taken so participants would believe the speech recognition system was real. The keyboard that was used by the experimenter was very quiet. It was placed on a soft base to reduce clicking noise and the experimenter pressed the keys very softly. The preset recognition errors that the system made consisted of acoustically confusable letters (e.g., T, P, and B: Teapody instead of Peabody). Words did not appear on the display immediately when the experimenter pressed the shortcut. Instead, they always appeared exactly 2 seconds after the last character was spelled, mimicking the delay of contemporary speech recognizers. Consistent with this approach, the address was accepted or rejected after exactly 3 seconds from the last digit.

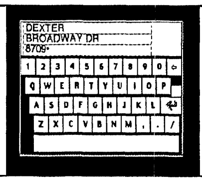
When the recognition system was in word dictation mode, the participant was instructed to say the address while slightly pausing between words. The experimenter used the same shortcuts that were used in the spelling mode. Accuracy per word was similar to spelling mode.

The Wizard of Oz technique, and the specific details of implementation in this study, proved to be very successful. All participants interacted with the system as they would interact with a real system. Most of them had no idea that the interface was simulated.

Control program

A custom SuperCard program (SuperCard ver. 3.6, IncWell Digital Media Group), running on a Power Macintosh 9500/120, was used to control the address entry task and record the participants' key presses. Table 5 shows the flow of a typical data-entry task using the touch-screen keyboard. The program started automatically when the driving simulator computer sent the experimental block to be performed over a connecting serial line. The program then displayed the first address of the specified block on the memo display (the small display to the right of the participant). One and a half seconds later, the address was spoken to the participant through speakers from prerecorded audio files. Four and a half seconds later, a beep cued the participant to begin entering the address into the simulated navigation system. At the same time, a signal was sent to the simulator to record the timing of the beginning of the trial in the driving performance data file. The SuperCard program then recorded the timing of every key press the participant made (at a resolution of 1/60 second). When the participant finished entering the destination, the program accepted the address if it was exactly the same as the instructed address or rejected it if there were errors. Fifteen seconds later, a new trial began. After six addresses were completed, or if the experimenter stopped the trials earlier, the program went into idle mode until the next block command was sent by the simulator.

Table 5. Flow of typical data-entry task using the touch-screen keyboard

Step	When	Example	Description of action
1	20 seconds after entering the road	"B"	The driving simulator computer sends a command to start a new block of 6 trials (6 addresses).
2	15 seconds later	DEXTER BROADWAY DR 8709	The in-vehicle task computer presents the first address on the memo display.
3	1.5 seconds later	"Dexter Broadway Dr Eighty seven zero nine"	The address is spoken through speakers behind the participant
4	6 seconds later	"Beep"	A beep cues the participant to begin entering the address.
5	When ready		The participant enters the address using the touch-screen keyboard, pressing Return after the city name, the street suffix, and the number.
6	After the final return is pressed		The experimenter compares the address entered to the correct address and accepts or rejects the message by displaying a corresponding message.
7	Option a	"Address rejected"	A "rejected" chime is played and "address rejected" is shown on the display.
	3 seconds after last keystroke of step 6		The participant uses the Backspace key to delete up to the position of the error and repeats the address entry from step 5.
	Option b	"Address accepted"	An "accepted" chime is played and "address accepted" is shown on the display.
	3 seconds after last keystroke of step 6		The next trial begins (step 2).

Address list

The addresses used in this experiment were carefully chosen to represent real locations. They were retrieved from an online address book (www.switchboard.com) using its national-search feature. Private addresses and business addresses were sampled from the databases of several states in different parts of the country to reduce the effects of geographical location and ethnicity on common street and city names. Half of the addresses were retrieved by searching for a typical business (Burger King restaurant). The rest of the addresses were retrieved by searching for individuals with certain common last names (Smith, Johnson, Anderson, Williams, Garcia). From the thousands of addresses found by the search engine, 300 were chosen at random. Certain criteria (Table 7) were then applied so that the addresses would meet the needs of this experiment.

Addresses were presented and entered in the order: city, street, street suffix, and building number, consistent with the order in which addresses are normally entered into navigation systems (Table 6). All addresses consisted of 20 characters.

Table 6. Sample addresses

City (characters)	Street (characters)	Street suffix (characters)	Number (characters)
Dexter (6)	Broadway (8)	Dr (2)	8709 (4)
Chicago (7)	Gannett (7)	Rd (2)	1672 (4)
Richmond (8)	Laurel (6)	St (2)	8564 (4)
Westfield (9)	Grand (5)	Dr (2)	2537 (4)

Table 7. Criteria for selecting addresses for the experiment

Purpose	Criterion
The difficulty of entering different addresses should be constant as much as possible to allow comparison across experimental blocks	All addresses were exactly 20 characters long.
	Names with difficult spelling or unexpected pronunciations were eliminated.
	All addresses were presented in the same order: a single-word city name, a single-word street name, and a 4-digit street number (direction prefixes such as North, East, etc. were dropped).
	Building numbers were chosen randomly (except for the first digit, which was always non-zero).
Reduce practice effects	Street names and city names did not repeat through the experiment (except in practice blocks).
	The number of characters in city names and street names varied (between 5 to 9).

Test Activities and Sequence

The participants began by completing a consent form (Appendix F) and a biographical form (Appendix G), followed by performing a vision test and a typing speed test. They were then seated in the driving simulator (Table 8– activity 1).

After a quick introduction to the study, the participants practiced driving for 5 minutes on a road that consisted of straight sections, moderate curves, and sharp curves (0, 3, and 9 degrees of curvature, respectively). They repeated a similar road in which baseline driving data were collected. In this block, and in later driving blocks, they were given a 1-dollar bonus if they remained in their lane for the entire time while maintaining the lead vehicles in sight.

The destination entry task was performed in blocks of 6 trials each (6 addresses). Keyboard entry blocks consisted of practice while parked, a baseline test while parked, practice while driving, 3 test blocks while driving, and a second baseline test while parked. Speech recognition entry blocks began in character spelling mode and continued in word dictation mode. The character spelling mode blocks consisted of practice spelling while parked, a baseline test while parked, and two blocks of spelling while driving. The word dictation mode blocks consisted of one test while parked and one test while driving.

After the destination entry tasks were completed, a second baseline road was driven. The participants then completed a post-test form (Appendix G) and a payment form, and received 40 dollars for their participation.

Table 8. Summary of activities and their sequence for a typical participant

Step	Task	Condition practice / test	Entry method	Driving workload *P=parked 0=low 3=moderate 9=high*	Estimated time (min)
1	Pre-test forms and tests				8
2	Practice driving	Practice		0, 3, 9	6
3	Baseline driving I	Test		0, 3, 9	6
4	Practice keyboard	Practice	Keyboard	P	8
5	Keyboard baseline I	Test	Keyboard	P	6
6	Practice keyboard and driving	Practice	Keyboard	0	7
7	Keyboard while driving	Test	Keyboard	0	7
8	Keyboard while driving	Test	Keyboard	3	7
9	Keyboard while driving	Test	Keyboard	9	7
10	Keyboard baseline II	Test	Keyboard	P	6
11	Break				7
12	Practice spelling	Practice	Spelling	P	8
13	Spelling baseline	Test	Spelling	P	7
14	Spelling while driving	Test	Spelling	0	7
15	Spelling while driving	Test	Spelling	9	7
17	Dictation baseline	Practice	Dictation	P	3
18	Dictation while driving	Test	Dictation	9	4
19	Baseline driving II	Test		0, 3, 9	6
20	Post-test debriefing				5
				Total	122

RESULTS

Overview

The destination-entry task completion time while driving, glance behavior, driving performance, and ratings of task difficulty are reported in this section, in that order. The destination entry section also includes predictions of task completion times based on task parameters.

What Factors Affected Destination Entry?

Effect of entry method on task completion time

Of the three entry methods examined, task completion time was shortest (14.7 s) when participants used speech recognition in word dictation mode (Figure 8). When using speech recognition in character spelling mode, task completion time was more than twice as long (39.1 s). The effect of curvature on task completion time in word dictation mode was negligible (14.2 s while parked and 15.3 s on sharp curves). The effect of driving-workload in character spelling mode was small (37.1 s while parked and 41 s on sharp curves, p=0.07), thus suggesting that speech input was relatively unaffected by the workload levels explored.

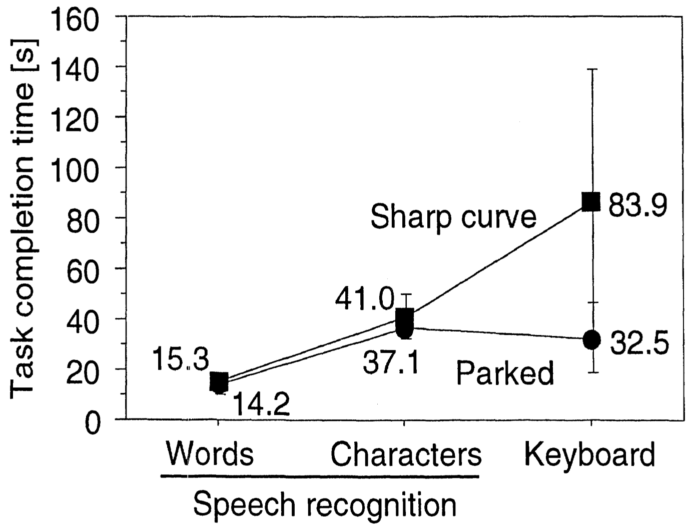


Figure 8. Task completion time by entry method

Although using the touch-screen keyboard while parked (32.5 s) was slightly shorter than in character spelling mode (p=0.07), on sharp curves it was much longer than either of the speech recognition modes (83.9 s, p<0.0001). On average, task completion time on a sharp curve was 2.58 times longer than when parked.

The effect of age on task performance was significant (p=0.0001) mainly due to the difference in touch-screen typing performance. Younger participants performed the typing task within 38 s while older participants needed more than twice the time (81.1 s). This finding is consistent with the literature where differences of 50-100% due to age are reported (Green, 2001c). The age effect in speech recognition mode was significantly smaller (p=0.0001; word dictation mode: 12.7, 16.8; character spelling mode: 36 s, 42 s, respectively).

Effect of curvature on keyboard task completion time

As pilot data had shown that the effects of curvature (workload) would likely only affect the keyboard task, a variety of workload levels were examined only for that task. As shown in Figure 9, task completion time increased significantly as a function of road curvature (p<0.0001), from 32.5 s while parked to 63 s on straight roads to 84 s on sharp curves (+94% and +158%, respectively). A Fisher's post-hoc test revealed that all pairs were different from each other at p<0.005 or lower, except for the difference between straight sections and moderate curves.

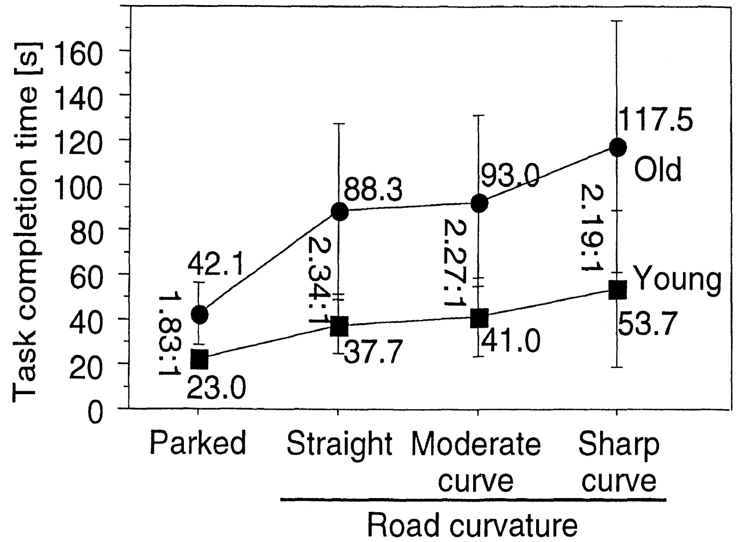


Figure 9. Keyboard task completion time by road curvature (workload) and age

Older participants took almost twice as long as younger participants to complete the task while parked (p=0.02) and more than twice as long while driving. The increase in task completion time due to the driving task (comparing straight to parked) was greater for the older participants (p=0.04). However, the increase due to road curvature (comparing a sharp curve to a straight road) was similar between the age groups.

Typing method (touch typist vs. hunt and peck) appeared to be important even when age-related differences were considered. As shown in Figure 10, within age groups, touch typists completed the task 25% faster than hunt-and-peck typists. However, typing speed on a 1-minute test, which was confounded with age and with typing method, was a better predictor of task completion time than typing method in this task. See Equation 4.

$$\text{Task completion time [s]} = 96 - 1.1 * \text{typing speed [words/min]}; \quad R^2 = 0.22 \tag{4}$$

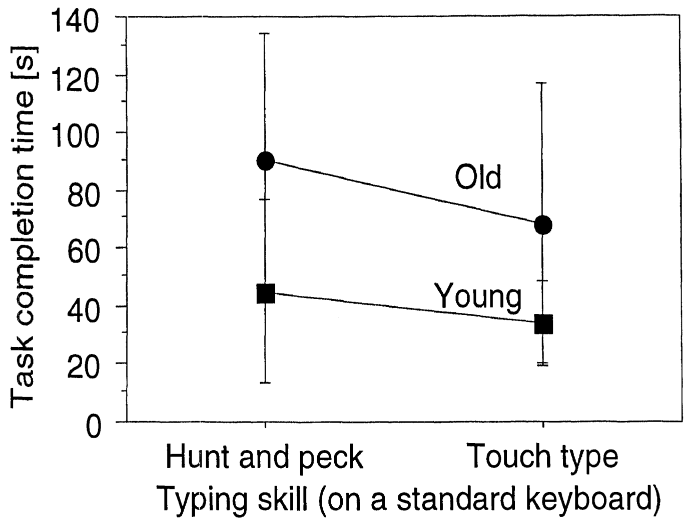


Figure 10. Task completion time by method of typing on a regular keyboard and age

Prediction of task completion time and error analysis

To gain further insight into the process of data entry while driving, a detailed analysis was performed on the entry keystroke data. Because the data reduction was extremely laborious, only a sample of 4 representative participants (1 from each age-gender group) was examined. (Refer to Appendix B for detailed analyses of all three destination-entry methods.)

To predict task completion time, equations were formulated for each of the mean measured word dictation, character spelling, and keyboard entry times, error correction times, and system recognition times. The predictions for each entry method were quite accurate. In particular, predictions for addresses with no errors came within +/- 2% of the actual entry times.

The experimental design allowed the prediction of entry times with certain numbers of errors because the error rate was controlled, which also allowed for predicting entry times with any given error rate.

Table 9 illustrates how detrimental errors are to the total task completion times. In the word dictation entry method, the act of detecting and correcting an error increased the task completion time by 71%.

Table 9. Predicted task completion times with zero and one errors

	Mean entry time with no error [s]	Mean entry time with 1 error [s]	Percent (%) difference
Word dictation entry	13.1	22.4	+71
Character spelling entry – parked	33.9	48.6	+43
Character spelling entry - driving	36.9	53.9	+46

Unlike in speech recognition mode, the experimenter had no control over the errors that participants committed in the keyboard-entry method. Errors were categorized into 3 groups: double clicks, keys adjacent to the correct key, and any other keys. For the 4 participants whose data were analyzed (1660 required key presses), the overall probability of committing an error was 4.7% per key, which amounted to about one error per address. The most common error was double clicking (2.6%), followed by pressing an adjacent key (1.3%) or any other keys (0.8%). On average, participants detected an error within 3.5 s of making it (range: 1 to 16 s).

Glance Behavior

Glance behavior was examined in detail only for the keyboard-entry task. As with the keystroke data, the laborious nature of the video data reduction process meant that only a subset of the data could be examined, in this case data from 12 participants, 3 from each age-gender group. Four addresses per participant were analyzed: two entered on a straight road and two on a sharp curve. The first 2 addresses were chosen because all participants had completed at least these two addresses. Glances are defined in many ways in the literature (Green, 2001a). The definition chosen here was consistent with SAE J2396 (Society for Automotive Engineers, 1999); that is, a glance began when the participant's eyes left the road and ended when the participant's eyes left the display or address –memo to look back to the road. Thus, a glance included the time to look at an in-vehicle display plus one transition.

Table 10 displays data for the 3 glance measures examined. (Note that video data reduction has an accuracy of plus or minus 0.1 seconds on all glance behavior measures.) For additional data, refer to Appendix D. Total glance duration, the sum of all individual glances combined, was 26 s for younger and 38 s for older participants. Mean glance durations were between 1.1 to 1.4 s depending on the participant's age and the road curvature. The mean time between glances increased significantly (from 1.0 to 1.4 s) with road curvature and age. The total number of glances per address was 20.6 for younger and 34.5 for older participants. Participants looked at the display more of the time on straight roads (60%) than on sharp curves (41%). Similarly, younger participants spent a greater proportion of their time looking at the in-vehicle display (63%) than did older participants (41%).

Table 10. Mean values of glance analysis measures

	Road curvature		Age	
	Straight	Sharp	Young	Old
Task completion time [s]	50.0	83.2	41.4	91.8
Total glance duration [s]	30.1	34.0	26.1	38.0
Glance duration [s]	1.4	1.1	1.4	1.1
Time between glances [s]	1.0	1.4	0.9	1.6
Number of glances	21.9	33.1	20.6	34.5

The effect of driving workload and age on total glance duration to the in-vehicle display

Total glance time to the in-vehicle display did not change significantly due to the driving workload level (p=0.28). There were, however, significant differences due to age (p=0.0001), with younger drivers requiring less glance time (Figure 11). The interaction between age and driving workload was not significant (p>0.05).

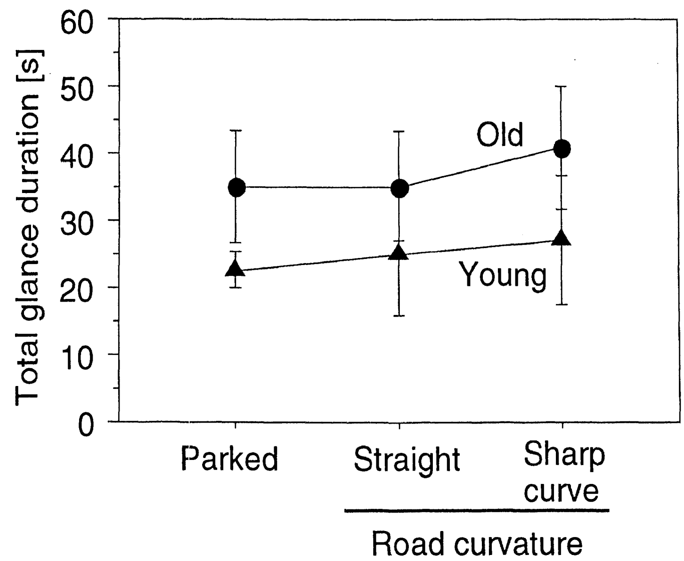


Figure 11. The effect of driving workload and age on total glance duration

The effect of curvature and age on glance duration and time between glances

Mean glance durations at the display decreased significantly ($p<0.0001$) as a function of road curvature and age (Figure 12). All participants made longer glances on the straight road (1.4 s) than on the sharp curve (1.1 s). The effect of age on glance duration was not significant ($p=0.10$), but the interaction between age and driving workload was significant ($p=0.04$). The decrease in mean glance duration was greater for younger participants.

Glances at the road scene (time between glances at the display) increased significantly as a function of road curvature and age (Figure 13). Participants made shorter glances at the road on the straight road (1.1 s) than on the sharp curve (1.4 s, $p=0.002$). Driving workload was higher on the sharp curve, which required participants to keep their eyes on the road for longer durations. Unlike glance duration at the display, the effect of age on time between glances was significant ($p=0.005$). Younger participants made shorter glances at the road (0.9 s) than did older ones (1.6 s).

Glance allocation and strategy

Figure 14 shows the relationship between the glance duration at the display and the time between those glances by participant. Classifying glance durations and time between glances into above and below a 1.0 second mean ("long" and "short"), 5 of the 6 younger participants made short glances to the roadway, while all of the older participants made long glances. However, younger participants did not make significantly longer glances to the display than did older ones, with equal numbers of young and older participants making short and long glance durations.

Figure 15 shows how participant glance behavior (or strategy) shifted from the straight road to the sharp curve. All participants made shorter glances at the display on the sharp curve. Most of the participants (11 of 12) made longer glances at the roadway. Both trends were expected, because of the higher visual demand associated with driving on the sharp curve.

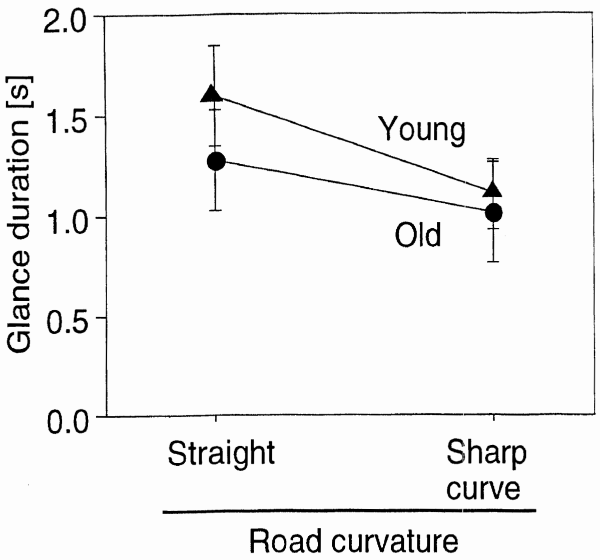


Figure 12. The effect of curvature and age on glance duration at the display

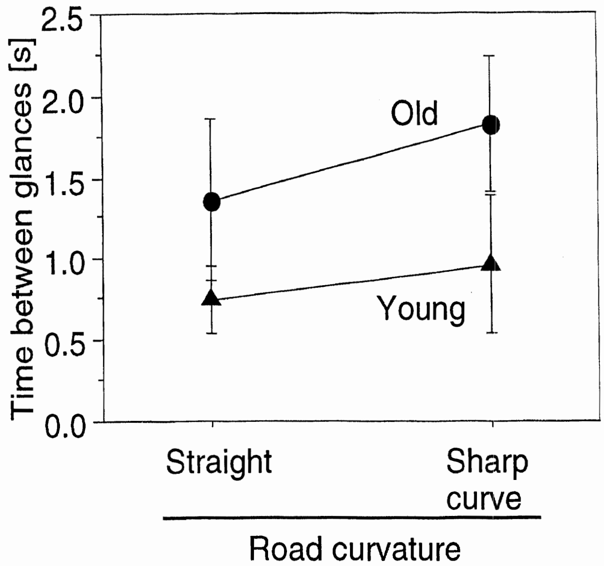


Figure 13. The effect of curvature and age on time between glances at the display

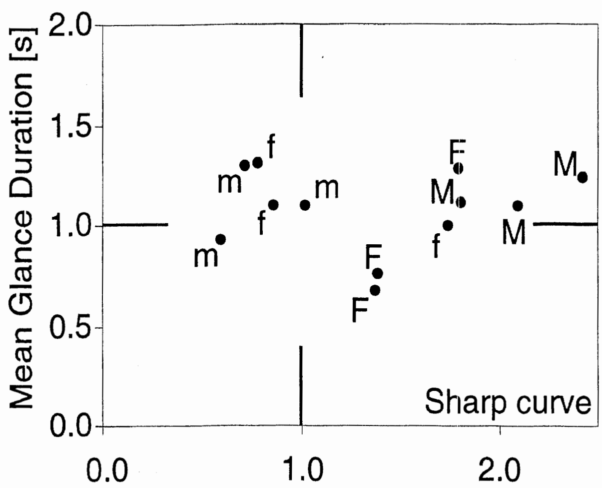


Figure 14. Glance allocation for sharp curve

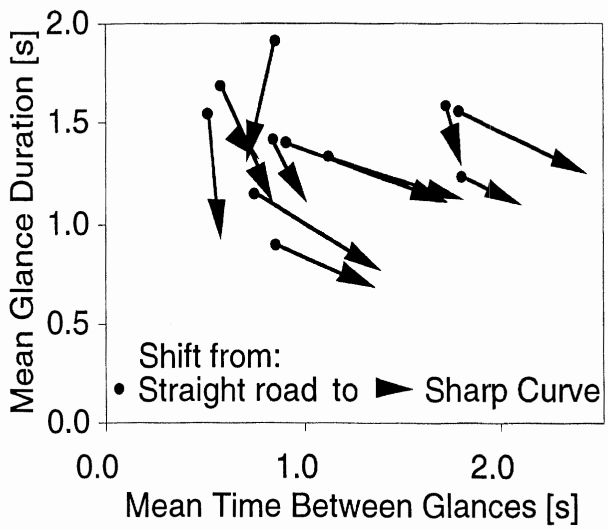


Figure 15. The effect of curvature on glance allocation

Comparison of glance behavior results to findings of 2 relevant experiments

Table 11 shows the comparison of glance behavior in this experiment to the previous experiment (Tsimhoni and Green, 2001) in which participants found street names and icons on a static map displayed in the vehicle. The task duration was much shorter than the current task, and did not require motor resources (i.e., using the finger to type). The total glance duration in the map task was about 5 s, versus about 32 s in the keyboard-entry task. While total glance time to the display decreased by 16% in the map task due to workload, there was an increase of 13% in the entry task (which was not statistically significant). Glance durations to the display in the map task were longer than in the entry task, but they decreased with road curvature in both tasks. Accordingly, the durations of glances to the road were longer in the map task, but they increased with road curvature in both tasks.

Although the durations of glances in each task were different, most of the trends were similar. When driving workload increased, glances to the display were shorter, glances back at the road were longer, and more glances were made. Total glance time decreased in the map task and increased in the entry task.

Table 11. Comparison of glance measures in this experiment to Tsimhoni and Green (2001)

	Previous experiment			Current experiment		
	Straight	Sharp	Difference	Straight	Sharp	Difference
Total glance duration [s]	5.0	4.2	-16%	30.1	34.0	+13%
Glance duration [s]	2.3	1.4	-39%	1.4	1.1	-21%
Time between glances [s]	1.1	1.8	+64%	1.0	1.4	+40%
Number of glances	2.6	3.5	+35%	21.9	33.1	+51%

Table 12 shows how results from Chiang, Brooks, and Weir (2000) compare to the results in the current experiment. Chiang et. al examined touch-screen destination entry while driving on the road. Ten Participants (ages 26 to 44, mean of 33) drove on city streets and the freeway while entering street addresses on an in-vehicle navigation system using a touch screen. The address entry task required about 15 keypresses, consisting of a few characters, 4 digits, and button presses for menu selection. They examined eye fixation times on the display and the number of keystrokes made within one fixation. (Due to the difference in definition between an eye fixation and a glance, a comparison of the results adds the duration of a single road-display transition to fixations, typically 0.15 s as reported by Chiang et. al.)

Although total glance duration was shorter in their experiment, the estimated glance times from Chiang et. al (1.15 s) compare well to glance duration in the current experiment (1.1 to 1.4 s). In addition, their finding that fixation times increased by 0.6 seconds with each additional keystroke is similar to the within-chunk times found in this experiment (given in the next section of the current experiment).

Table 12. Comparison of glance measures to Chiang, Brooks, and Weir (2000)

Chiang et. al (2000)		Current experiment		
Description		Description	Straight	Sharp
Total glance duration	About 18	Total glance duration	30.1	34.0
Fixation time [s] Estimated glance duration	1.0 1.15	Glance duration [s]	1.4	1.1
Fixation time per additional keystroke [s]	0.6	Keying time per character [s]	0.8	0.6

Task Partitioning (Character Entry)

Recent discussions of automotive safety standards, in particular SAE J2364 (Society of Automotive Engineers, 2000) have raised questions about how people partition tasks when driving. For that reason, pausing between fields of text and groups was examined. (See Figure 16.) The time between fields was defined as the interval between the entry of a field-delimiting character (e.g., carriage return or space) and the entry of the next appropriate character. A group was defined as a sequence of characters that were entered during a single glance.

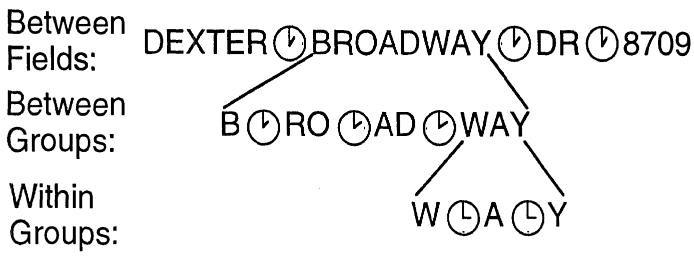


Figure 16. Breakdown of character entry measures

It was expected that when entering addresses the participants would make longer pauses between fields where a pause would occur naturally. The 3 field delimiters examined were the return after the city name, the space after the street name, and the return after the street suffix. Within fields, characters would not necessarily be entered individually, but rather in groups of characters. The time between groups was defined as the duration between the last character of a group and the first character of the following group. The time within groups was the duration between the characters that composed a group. The time between groups was expected to increase as a function of curvature, whereas the time within groups was expected to remain constant. The data to validate these expectations appear in Table 13. For additional data refer to Appendix D.

Table 13. Mean values of task partitioning measures

	Road curvature			Age		
	Straight	Sharp	Difference	Young	Old	Difference
Time between fields [s]	3.5	5.5	+57%	2.8	6.2	+121%
Time between groups [s]	2.6	3.3	+27%	2.1	3.7	+76%
Time within groups [s]	0.8	0.6	-25%	0.6	0.8	+33%

The Effect of curvature and age on pauses between fields and groups

Given the effects of visual demand (curvature) on the time between glances just described, where did the changes occur? As shown in Figure 17, pauses between fields on the straight road (3.5 s) were significantly shorter than on the sharp curve (5.5 s, p=0.01). Younger participants had significantly shorter times between fields (2.8 s) than did older ones (6.2 s, p=0.005).

Likewise, the time between groups increased significantly as a function of road curvature and age (Figure 17). The participants made shorter pauses between groups (2.6 s) on the straight road than on the sharp curve (3.3 s, p=0.0009). Younger participants had less time between groups (2.1 s) than did older participants (3.7 s, p=0.01).

Given that glance times decreased with visual demand, and that there were substantial increases between fields and groups, little change within groups was expected. The time within groups could only be calculated when more than one character was entered in a single glance. Figure 17 shows the glance times with double-click errors removed. (Not all participants entered multiple characters in a single glance, so the average between address 1 and 2 for each participant was used for analysis of variance.) The participants had longer time between characters within groups (0.8 s) on the straight road than on the sharp curve (0.6 s, p=0.06). Consistent with the other measures, younger participants spent less time (0.6 s) between characters within groups than did older participants (0.8 s, p=0.02).

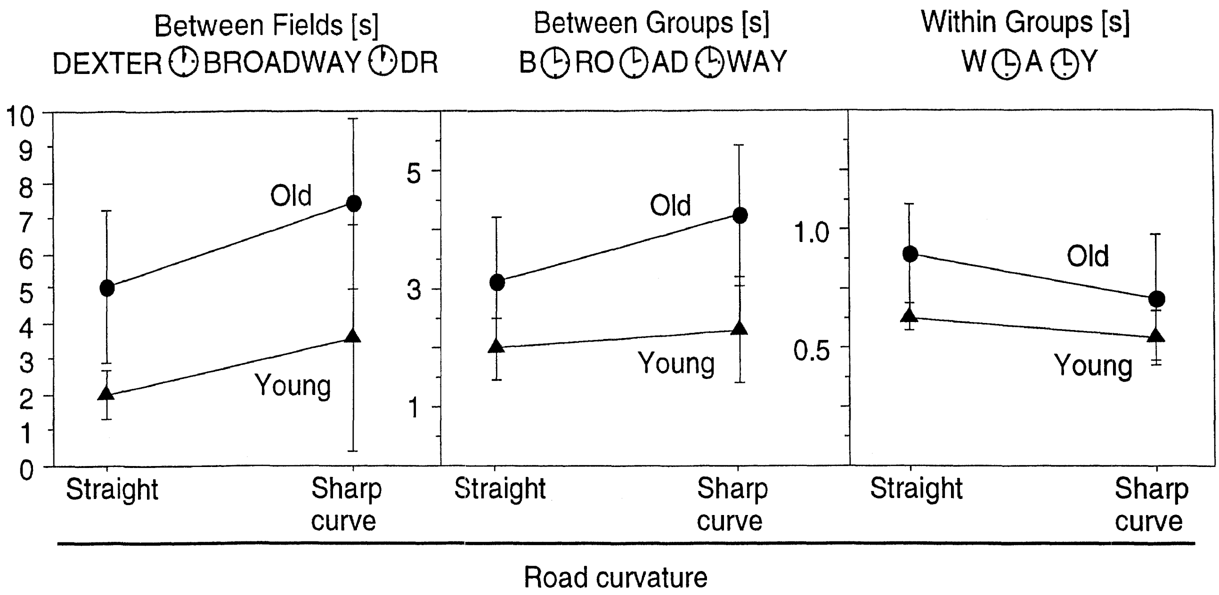


Figure 17. The effect of curvature and age on task partitioning

Driving Performance

Seven measures of lateral and longitudinal control of the vehicle were used to quantify the driving performance of participants. Table 14 provides a brief definition of each measure of driving performance and how it was analyzed. The data for most measures were recorded by the simulator at 30 samples per second.

For many of the measures examined, there are no standard definitions, which complicates comparison with other studies in the literature. For example, a lane departure could be registered when any tire of the participant vehicle touches a lane marking, when it crosses into another lane, when a part of the vehicle (e.g., an outside mirror) protrudes over a lane marking, or as any one of several other possibilities.

Table 14. Lateral and longitudinal driving performance measures

	Measure	Analysis[2]	Definition
Lateral	Lane departures	Number of events	Vehicle leaves roadway or enters adjacent driving lane[1]
Lateral	Lateral lane position [m(ft)]	Mean, Standard deviation	Lateral position of center of vehicle in driving lane
Lateral	Yaw angle [radians]	Mean, Standard deviation	Angular heading of vehicle (0 degrees is straight ahead; positive angles are to the right)
Lateral	Steering wheel angle [radians]	Standard deviation	Angle of steering wheel (0 degrees steers the vehicle straight ahead; positive values are to the right)
Longitudinal	Headway [m(ft)]	Mean, Standard deviation	Distance from front of participant's car to front of lead vehicle
Longitudinal	Forward velocity [km/hr(mi/hr)]	Mean, Standard deviation	Forward traveling velocity of participant's vehicle
Longitudinal	Accelerator position [%]	Standard deviation	Percent throttle applied to participant's vehicle

Note 1. A unique lane departure was registered when one of the tires touched a lane marking after the vehicle had been within the lane markings for at least 1 s. If the vehicle left the lane, returned for an interval shorter than 1 s and left the lane again, it was considered a single departure.

Note 2. The mean and standard deviations of each driving performance measure were calculated in 5-seconds bins, which allowed comparison between tasks of different durations. The values reported (the mean within a 5-second bin and the standard deviation within a 5-second bin) were averaged across participants and across bins.

Figure 18 presents typical driving performance of a young woman entering addresses in 3 conditions: no task, speech recognition-word dictation mode, and touch-screen keyboard entry. The variability of the driving performance measures is lowest for the no-task condition, and highest during touch-screen keyboard entry. Also shown is how over time, forward velocity slows below the suggested speed of 66 km/hr [45 mi/hr]. Finally, it can be seen that headway generally increases over time, and increases among the 3 conditions.

Appendix J contains an explanation for the observed artifact of high-frequency noise in the yaw angle and lateral position plots in Figure 18.

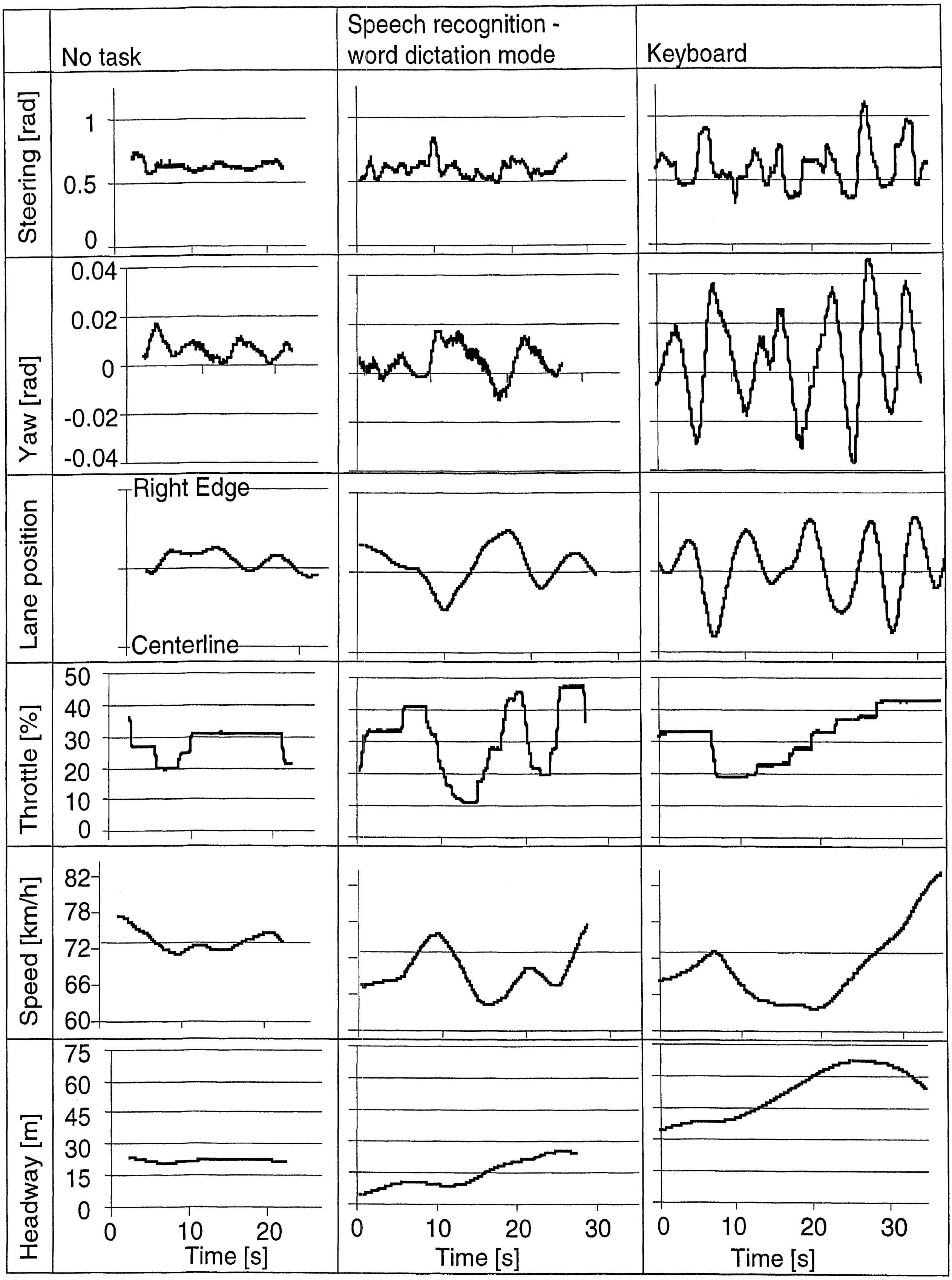


Figure 18. Typical driving performance for a destination entry trial

The effect of entry method on driving performance

The driving performance on the first 10 seconds (2 bins of 5 seconds each) of 2 addresses on a sharp curve was analyzed for each of 3 entry methods. Segments of similar length and curvature were recorded for the no-task condition as baseline. This approach was chosen to provide a common data set for analysis. The alternative, of using mean trial performance, would have been less accurate because the task completion time for each trial was different. However, the choice of this approach will lead to analyses that may underestimate timing between task differences.

Lateral control

Degradation of lateral control has significant safety implications. Analysis of 4 measures of lateral control (steering, yaw angle, lateral position, and lane departures) showed that when entering an address using a keyboard, driving performance was significantly worse than in the other conditions, but there was no difference between performance in the baseline and speech recognition conditions.

The mean steering wheel angle did not change with entry method (0.6 rad, 34 degrees in all conditions). However, the standard deviation of steering wheel angle when using the keyboard (0.12) was 60% higher than in the other conditions (no task=0.070, word dictation mode=0.085, character spelling=0.080; $p<0.0001$) (Figure 19).

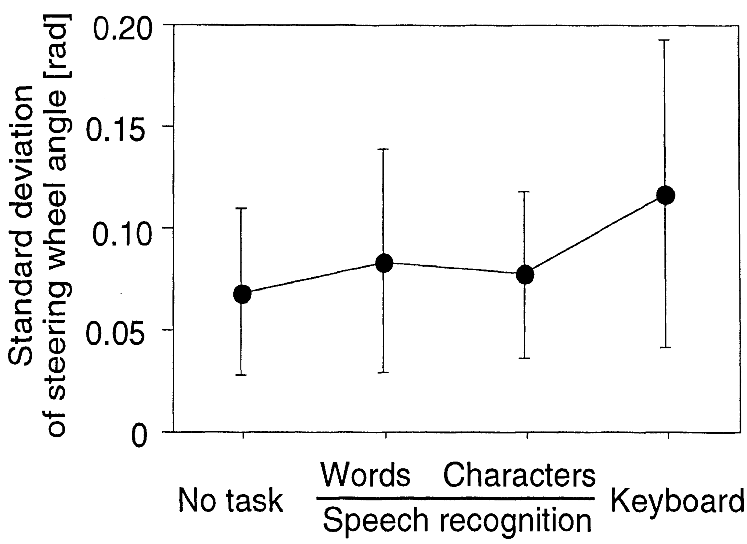


Figure 19. The effect of entry method on standard deviation of steering wheel angle

The mean yaw angle did not change with entry method. The standard deviation increased from 0.015 in the speech condition to 0.020 in the keyboard condition ($p<0.0001$). The mean lateral position followed the same pattern. The standard deviation of lateral position increased 60% from 0.13 to 0.21 m in the keyboard condition ($p<0.0001$) (Figure 20).

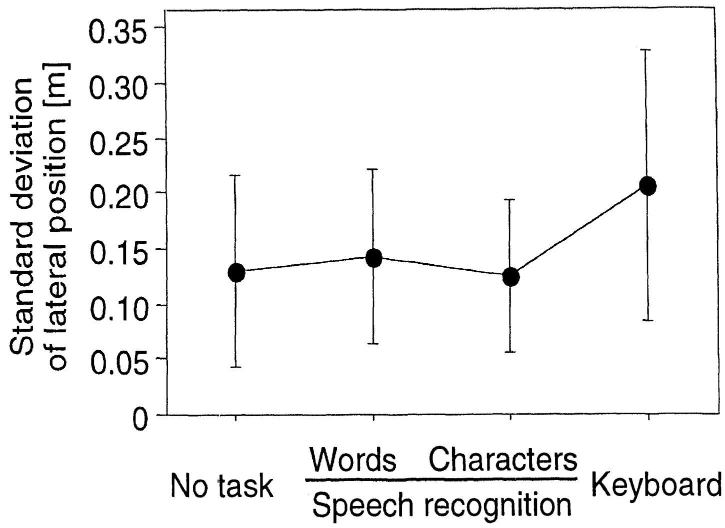


Figure 20. The effect of entry method on standard deviation of lateral position

Lane departures

The main measure of interest was the percentage of addresses in which at least one lane departure occurred. This measure was preferred over the total number of departures for a number of reasons. In real-life, a first lane departure might be the last one, so its significance is greater than potential consecutive departures. In an experimental setting, and particularly in a simulator, a driver is more likely to commit multiple lane departures because they cause no physical consequences to the driver or the vehicle. Finally, while some of the participants responded to their initial departure by being more careful (at least for a while), others did not. This difficulty might have led to an inflated number of lane departures, thus providing misleading results.

When using the keyboard to enter an address, one or more lane departures occurred while entering 57 out of 276 address (20.6%). With the speech recognition method in character spelling mode, the number of departures was 4.2% (11 out of 263 addresses) and in word dictation mode 8.3% (12 out of 144 addresses). (Note that the latter is an overestimation because word mode was tested on sharp curves only.) Overall, 169 lane departures were committed by all the participants throughout the experiment. In 79 out of 683 addresses (11.6%), one or more departures occurred while typing the address. In contrast, only 1.5% of the no-task road segments (2 out of 131) had lane departures.

An additional measure of interest was the departure rate per minute. It provides an estimation of the difficulty of the task, independent of the duration required for completion. (For a detailed discussion, see Nowakowski, Utsui, and Green, 2000). The mean number of lane departures per minute was calculated across all participants (total number of lane departures divided by total task time across all participants). When using the keyboard, the lane departure rate was 0.51 per min overall and 0.88 on sharp curves. In character spelling mode it was 0.10 overall and 0.19 on sharp curves. In word dictation mode it was 0.37 on sharp curves (other roads were not tested). The driving baseline was 0.01 overall and 0.08 on sharp curves. The departure rate for word mode was high relative to spelling mode when considered in the context of the duration of the task (the rate was twice is large), but in practice, the ratio between the number of departures was much lower because word dictation was relatively fast.

Table 15. Lane departure rate by entry method and age

- Number of addresses with ≥ 1 departure
- Probability of ≥ 1 departure per address
- Rate of departures per minute

	Young	Old	Sharp curve	Total
No task	0/66 0% 0	2/65 3.1% 0.02	1/24 4.2% 0.08	2/131 1.5% 0.01
Word dictation	1/72 1.4% 0.11	11/72 15.3% 0.58	12/144 8.3% 0.37	
Character spelling	0/138 0% 0	11/125 8.8% 0.19	10/133 7.5% 0.19	11/263 4.2% 0.10
Keyboard	21/182 11.5% 0.27	36/94 38.3% 0.74	27/80 33.7% 0.88	57/276 20.6% 0.51

Normally in experiments, the lateral control measure of choice would be standard deviation of lane position, because of its sensitivity. In this experiment, lane departures were more sensitive to differences in entry method, though not all participants committed lane departures. Despite not having greater standard deviations of lateral position, older participants committed most of the lane departures (Table 16).

Table 16. Number of lane departures by age and gender

	Young female	Young male	Old female	Old male
Number of participants who committed at least one lane departure	5	2	6	6
Total lane departures	20	14	45	78

Figure 21 shows the time distribution of first departures for the keyboard-entry task. Most, but not all, first departures occurred within the first minute of performing the data-entry task.

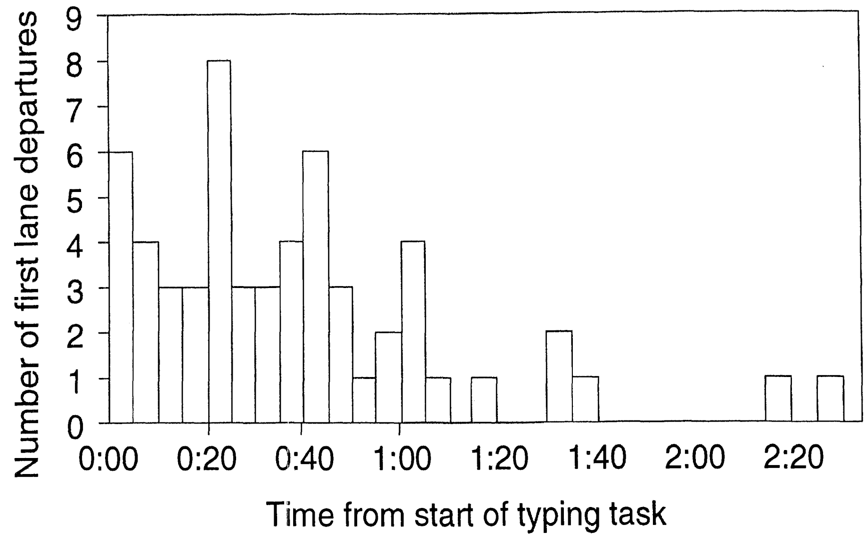


Figure 21. Timing of first lane departures while typing an address

Figure 22 shows the rate of lane departures as a function of task completion time. While older participants showed a slight increase in the rate of lane departures as task completion times were longer, younger participants did not. There was no clear cut-off point at which the lane departure rate increased significantly. In some instances, task completion times over 150 seconds yielded no departures at all.

However, there is no increase in departure rate when considering departure rates and task times within participants. Consequently, the trend here may just be the association between the difficulty of the overall task for different participants and the corresponding task duration and lane departure rate.

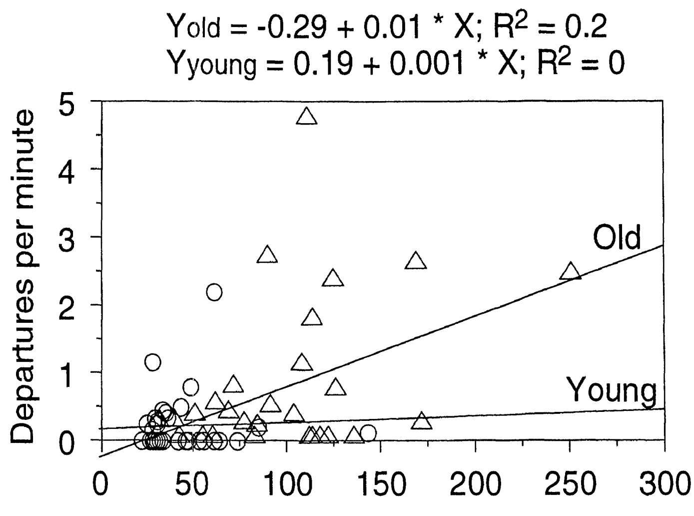


Figure 22. The effect of task completion time on lane departure rate

Change in lateral control as a function of time

The standard deviation of lateral position while using the keyboard varied throughout the entry process, which suggests a cumulative effect on driving performance. To avoid bias for the longer tasks, only the first 25 seconds of each task were analyzed; in some cases this was the length of the full task, in others it was only the beginning.

The standard deviation of lateral position increased in the first 25 seconds ($p<0.0005$). On sharp curves, it increased from 0.18 m in the first 5 seconds to 0.27 m after 20 seconds. On moderate curves, the effect was smaller (from 0.11 to 0.17 m). On straight sections, there was a slight increase after 15 seconds but the value at 25 seconds was similar to that at 5 seconds.

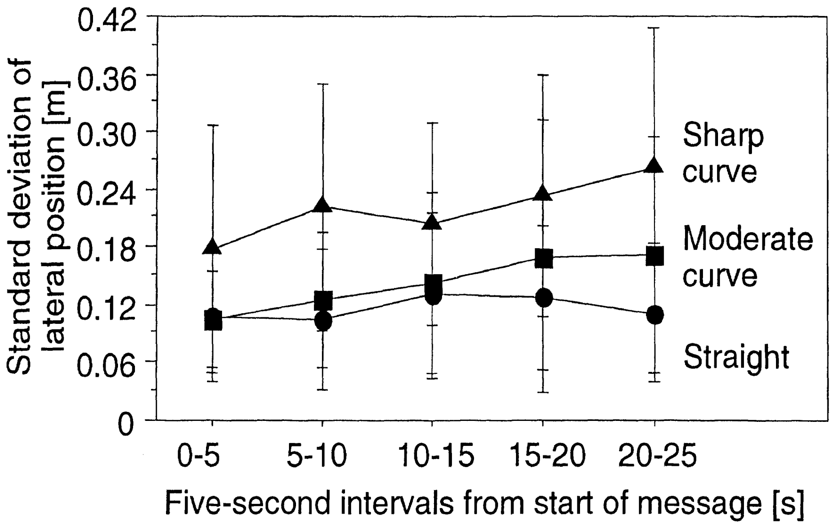


Figure 23. Standard deviation of lateral position by workload and time

<u>*Longitudinal control*</u>

Analysis of 3 measures of longitudinal control (accelerator position, forward velocity, and headway) showed a mixed pattern. When entering an address, regardless of the entry method, the mean forward velocity dropped. The following distance was least in the word dictation mode and highest when using the keyboard. The change in following distance may be a partial compensation mechanism to provide increased driving safety.

The mean accelerator position in character spelling mode (31.5%) was higher than all other conditions (28%, $p<0.05$). The standard deviation of accelerator position was high when no task was performed (6%), moderate in character spelling mode (5%), and low in the other two modes (4%) ($p<0.005$).

Since the lead vehicle in the experiment traveled at 72 km/hr, participants were instructed to follow at a comfortable distance, also at 72 km/hr. The mean forward velocity while entering addresses in any of the methods (70.2 km/hr) was somewhat lower than just driving (73.4 km/hr, $p<0.01$) (Figure 24). In character spelling mode, the forward velocity was slightly lower than in keyboard mode ($p<0.1$). The standard deviation of forward velocity did not vary significantly between the methods (1.5 km/hr). For all entry methods, it could be hypothesized that when the task began, forward velocity would decrease from the suggested 72 km/hr until the task ended, at which point the participants would increase again to the suggested forward velocity. However, increases and decreases in forward velocity happened at all points during an entry task, and varied by participant and task duration.

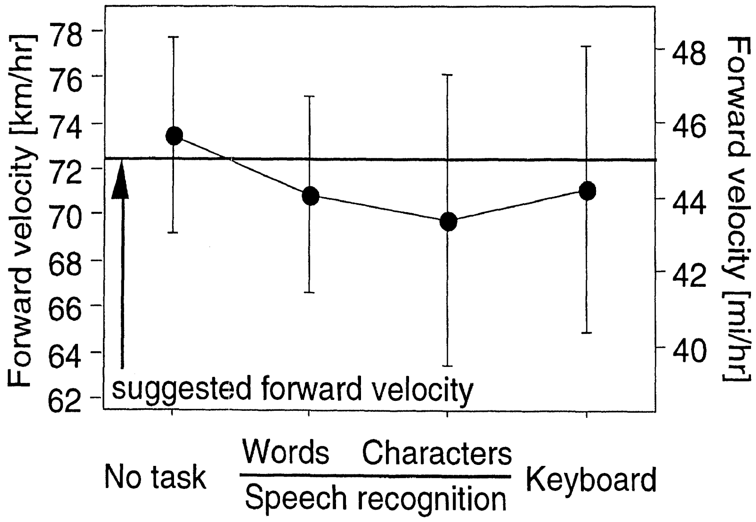


Figure 24. The effect of entry method on mean forward velocity

The following distance was shortest in word dictation mode (88 m) and longest when using the keyboard (167 m; $p<0.05$) (Figure 25). The standard deviation of the following distance with no task and in word dictation mode (4.9 m and 4.6 m, respectively) was lower than in character spelling mode and when using the keyboard (6.5 m and 6.1 m, respectively).

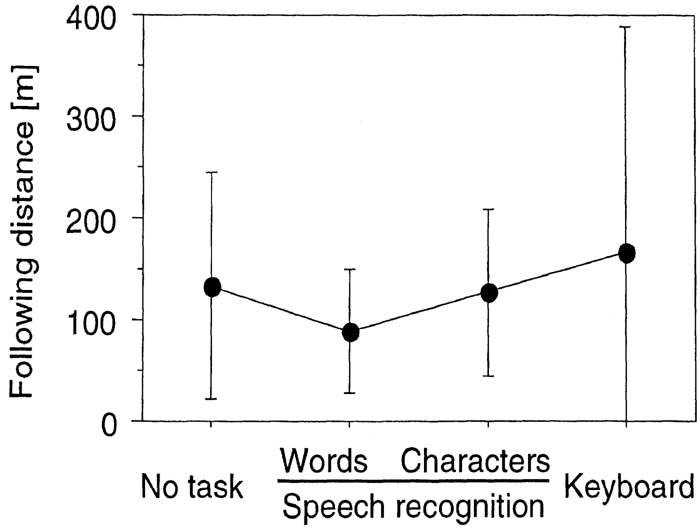


Figure 25. The effect of entry method on following distance

Subjective Ratings

Difficulty

After each block, the participants rated the task difficulty and the difficulty of driving while performing the task using a modified Cooper-Harper Scale (Appendix G). For task difficulty, the effect of entry method, the effect of driving, and the interaction between them were significant (p=0.0001) (Figure 26). While driving baseline was rated at 2.2, driving while typing on the touch-screen keyboard was rated at 5.4. A Fisher's post-hoc test showed that touch-screen keyboard entry was rated significantly more difficult than any of the other 3 conditions (p=0.0001). While no difference was found between the driving baseline and the speech recognition conditions, spelling mode was more difficult than word mode (p=0.03). Data entry while parked was similar for all entry methods.

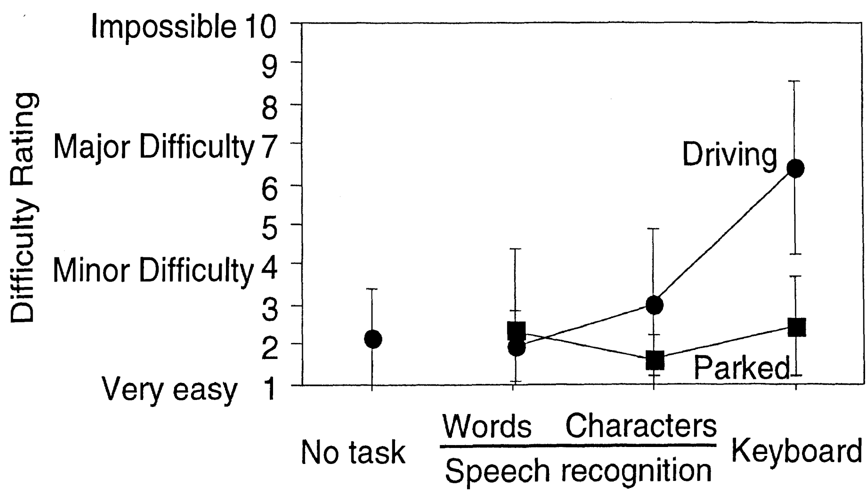


Figure 26. Difficulty ratings by entry methods while parked and while driving

Age differences were not statistically significant (p=0.33), although older participants tended to rate the keyboard task as more difficult than did younger participants (7.1 versus 5.7, respectively).

Figure 27 shows the effect of road curvature on the subjective rating in each of the entry conditions. Typing the addresses on a keyboard in any of the driving workload levels was significantly more difficult than while parked (p=0.0001). In addition, a Fisher's post-hoc test showed that the difficulty on the sharp curve (7.3, moderately difficult) was different from on the straight section (5.5, midway between easy and impossible) (p=0.01), but not significantly different from the moderate curve (6.0) (p=0.013).

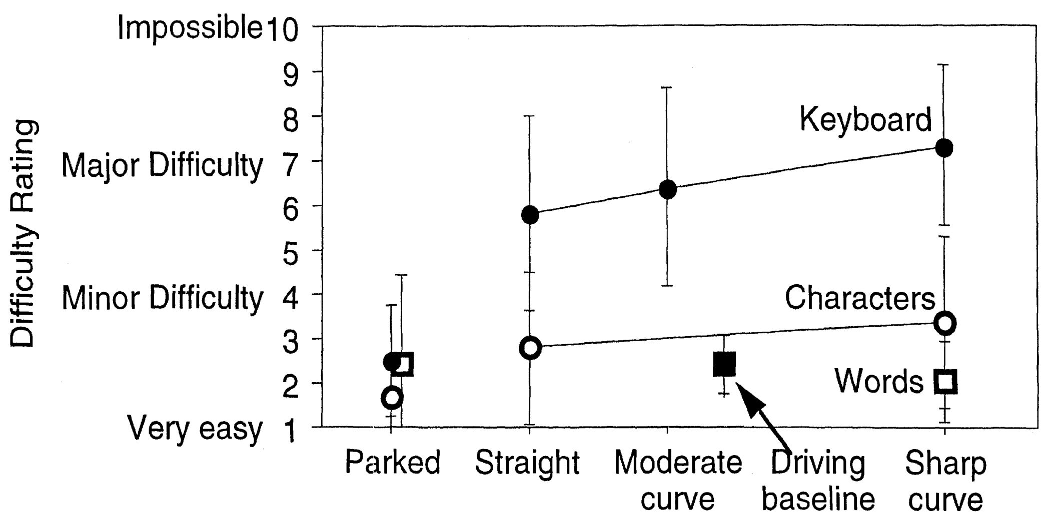


Figure 27. The effect of driving workload on difficulty ratings

Note: For driving baseline, a road consisting of several curvatures was rated. It is therefore displayed at the mean value of the curvatures tested.

Subjective rating of spelling the addresses by character using speech recognition followed the same pattern, significantly increasing with driving workload (p=0.0001). A Fisher's post-hoc test showed that the parked condition (1.7) was significantly different from the straight (2.8) and the sharp curve (3.4), but the 2 driving conditions were not different from each other.

Entering the addresses in word dictation mode was not affected by driving workload (p=0.66). There was a slight non-significant decrease in difficulty between the parked condition (2.4) and the sharp curve (2.1), which is attributed to confounding with task order. (In fact, the word dictation mode was rated slightly easier than the driving baseline even though it was on a sharp curve, while the driving baseline rating included all curvature levels).

Post-test evaluation

In addition to ratings at the end of each block, participants also rated the safety and difficulty of the different entry methods at the end of the experiment.

The rating of difficulty followed a pattern similar to the rating given after each block. The effect of driving, the effect of entry method, and the interaction between them were significant (p=0.0001). As seen in Figure 28, keyboard entry while driving (6.9) was rated as more difficult than the speech recognition modes (character spelling: 4.3 and word dictation: 2.9). Keyboard entry while parked (2.6) was not rated significantly higher than the speech recognition modes (character spelling: 2.1 and word dictation: 1.5), but a linear trend existed. This linear increase, however, did not appear in the subjective ratings given after each block. It is most likely that the difficulty rating while parked was affected by the participant's perception of the difficulty while driving. Thus, keyboard entry while parked was rated slightly more difficult than speech recognition, whereas in the rating collected after each block they were rated as equally difficult.

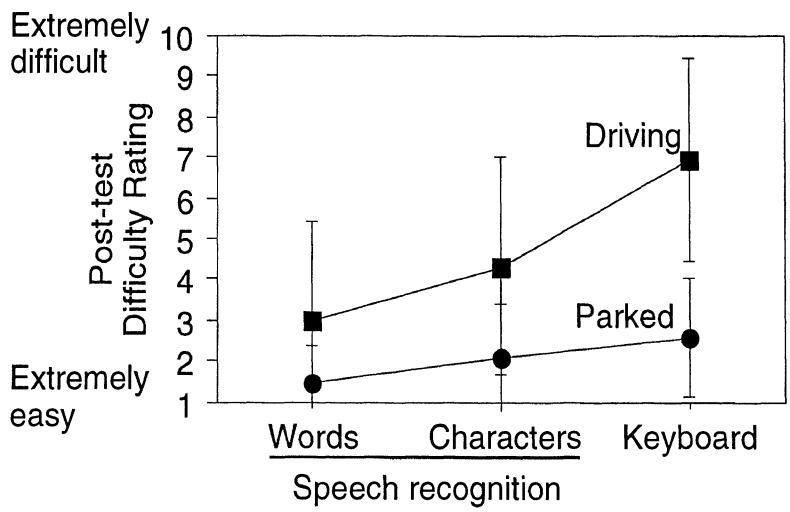


Figure 28. Post-test evaluation of difficulty by entry method

Participants were asked how unsafe each of the address entry methods was. The entry method was significant (p=0.0001). Speech recognition in word dictation mode was rated relatively safe (4.1), followed by character spelling mode (5.3). Keyboard entry was rated very unsafe (9.2). Ten out of 12 older drivers rated the keyboard 10 (very unsafe), while only 4 younger drivers rated it 10. No significant differences were found between younger and older drivers, or between men and women (Table 17 and Figure 29). This is a very strong indictment of keyboard entry while driving.

Table 17. Distribution of rating for the keyboard task by age

		"It is unsafe for me to enter addresses while driving" (number of participants for each rating from 1 to 10) (1 = strongly agree, 10 = strongly disagree)									
		1	2	3	4	5	6	7	8	9	10
Word dictation	Young	2	5	3		1					1
	Old	3	2	2					2		3
Character spelling	Young		1	3	2	3	1	1			1
	Old	1	2	2	2						5
Keyboard	Young				1			2	4	1	4
	Old								2		10

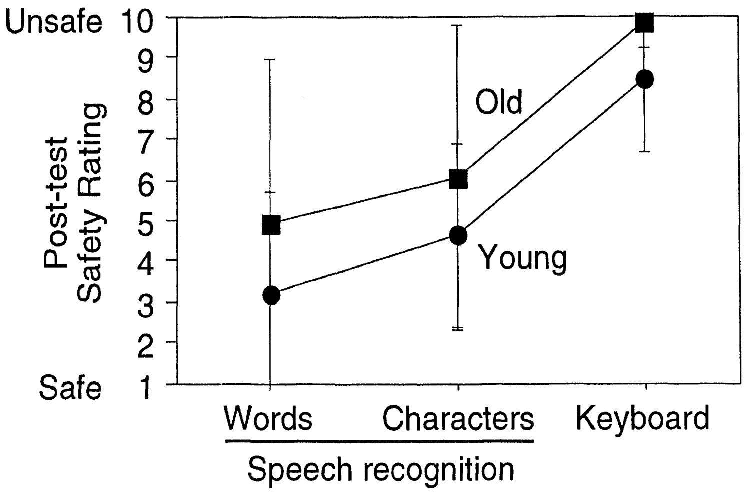


Figure 29. Post-test evaluation of safety by entry method

Summary

Figures 30 and 31 display a summary of the results and significant differences from the destination entry and driving performance measures. An area that is shaded gray indicates that particular entry method or road curvature was not examined for the corresponding measure. The arrows indicate whether the difference between two measures was significant. A horizontal arrow indicates no change, an arrow sloping upwards indicates a significant increase, and an arrow sloping downwards indicates a significant decrease.

	Driving workload			
	Parked	Straight	Moderate	Sharp
Destination Entry Task Completion Time				
Speech recognition - word dictation mode [s]	14.2			15.3
Speech recognition - character spelling mode [s]	37.1	41.1		41.0
Touch-screen keyboard [s]	32.5	63.0	67.0	83.9
Glance Behavior				
Total glance duration [s]	28.8	30.1		34.0
Glance duration [s]		1.4		1.1
Between glance duration [s]		1.0		1.4
Number of glances		21.9		33.1
Task Partitioning				
Between fields [s]	1.6	3.5		5.5
Between groups [s]	1.0	2.6		3.3
Within groups [s]		0.8		0.6

Figure 30. Summary of destination entry-related results

	Entry method			
	No task	**Word**	**Spelling**	**Keyboard**
Driving Performance				
Lane departures/minute	0.01	0.37	0.10	0.51
Lateral position [m]	0.13			0.21
Yaw angle [m]	0.015			0.020
Steering wheel angle [rad]	0.070			0.120
Headway [m]	133	88	129	167
Forward velocity [km/hr]	73.4	70.2		

Figure 31. Summary of driving performance results

SUMMARY AND DISCUSSION

Issues Addressed

1. *How does the total time for drivers to enter destinations using manual and voice methods compare as a function of the driving workload? (Which is best and when?)*

Task completion time while parked was shortest when using speech recognition in word dictation mode (14.2 s), followed by the keyboard-entry method (32.5 s) and speech recognition in character spelling mode (37.1 s). On sharp curves, task completion time with the keyboard increased significantly (+158%; 83.9 s). In contrast, speech recognition completion times increased only slightly in character spelling mode (+11%; 41 s) and in word dictation mode (+8%; 15.3 s).

The time to detect and correct errors was a substantial part of the destination entry time. Errors added about 9 s (+71%) to the task completion time of a single address in word dictation mode, about 17 s (+46%) in character spelling mode, and about 11 s (+20%) when using the keyboard.

As expected, destination entry in word dictation mode was the quickest method, followed by character spelling and keyboard entry. However, it was not expected that keyboard entry would be faster than character spelling mode while parked. The effect of moderate levels of driving workload on destination entry was nonsignificant when using speech recognition in word dictation mode, limited in magnitude in character spelling mode, and extremely significant when using the keyboard.

2. *When using a touch-screen keyboard to enter addresses, how do the inter–item intervals (mean time and variability) vary as a function of the driving workload? How does the pattern of pauses differ from the single task (nondriving) situation?*

Pauses between fields (e.g., between the city name and the street name) increased significantly with road curvature, from 3.5 s on straight sections to 5.5 s on sharp curves (+57%). Likewise, pauses between groups of characters increased from 2.6 s on straight sections to 3.3 s on sharp curves (+27%). Within groups, the time between characters decreased from 0.8 to 0.6 s on sharp curves (-25%).

A comparison of this pattern of pauses to the single task (nondriving) condition revealed a similar trend. The pauses between fields increased due to driving (from 1.6 s while parked to 3.5 s on a straight section; +119%) and the time between characters decreased as a function of driving (from 1.0 s while parked to 0.8 s on a straight section; -20%).

Participants compensated for higher workload by making longer pauses between fields and between groups of characters. Moreover, the number of consecutive characters in each glance was smaller and therefore the time between consecutive characters was shorter. This conforms to expectations.

3. *What is the pattern of glances for destination entry using a touch-screen keyboard and how does it change as a function of driving workload?*

Glances to the in-vehicle display on sharp curves were shorter (-21%) and the intervals of glances back at the road were longer (+40%) than on a straight road. The number of glances increased by 51%, but total glance duration (the total time spent looking at the display) did not change significantly (a nonsignificant increase of +13%).

The analysis of glances suggests that task completion time increases with driving workload primarily due to longer glances back to the road between glances at the in-vehicle display. In addition, as glances at the display become shorter, more glances are made. However, the cost of partitioning the keyboard task in this experiment into shorter elements was not high so only a slight increase in total time spent away from the road occurred.

The change in the pattern of glances was similar to the previous experiment (Tsimhoni, Yoo, and Green, 1999), in which tasks were only visual and significantly shorter. However, while the previous experiment saw a slight decrease in total glance duration, the current experiment saw a slight increase.

4. *How much does the addition of manual and speech input tasks degrade driving performance for each level of workload?*

Lateral vehicle control when entering an address using a keyboard (measured by the standard deviations of lateral position and steering wheel angle) was significantly worse than the baseline no-task condition (+60%). The probability of a lane departure while entering an address was 20.6% relative to 1.5% in the baseline condition, a slightly higher than 13-fold increase. There was no significant difference in most measures (variation in lateral position, yaw angle, and steering wheel angle) between performance in the speech recognition conditions and in the baseline condition. The probability of lane departures in sharp curves, however, increased from only 4.2% in the no-task baseline to 7.5% in character spelling mode and 8.3% in word dictation mode.

Longitudinal vehicle control presented a different pattern. When entering an address, using any of the entry methods, the mean speed dropped (by a mean of 1.8 km/hr). The mean following distance was lowest in the word dictation mode (88 m) and highest when using the keyboard (167 m).

Driving performance degradation was detected in all entry methods despite the simplicity of the driving task. Lane departures, the most notable consequence of not paying attention to the road, occurred mostly while using the keyboard but also when using speech recognition.

5. *How did participants rate the difficulty and safety of the tested entry methods?*

During the experiment, participants gave similar ratings to the difficulty of the 3 entry methods while parked (2.2 on a 10-point scale). While driving, the keyboard-entry method was rated most difficult (5.4), followed by character spelling (3.0) and word dictation (2.3). In the post-test questionnaire, the keyboard task was rated slightly more difficult than the speech recognition tasks in the parked condition (2.6 versus 2.1 and 1.5, respectively) and significantly more difficult in the driving conditions (6.9 versus 4.3 and 2.9, respectively).

In post-test safety ratings, the keyboard task was rated extremely unsafe (9.2) while the two speech recognition tasks were rated neutral (4.1 and 5.3).

Most participants agreed that using a keyboard while driving was extremely unsafe. However, they were not as consistent in their perception of safety for speech recognition destination entry. While some rated it extremely unsafe, some rated it, especially in word dictation mode, close to safe or extremely safe.

6. *How does entry performance and behavior vary as a function of driver age and gender?*

Older participants took nearly twice as long as younger participants to complete the keyboard-entry task while parked (+83%) and more than twice as long while driving (+127%). When using speech recognition, age differences were less prominent (+17% and +32% for word dictation mode and character spelling mode, respectively).

A detailed analysis of data from a few participants revealed that older participants typed each key more slowly. In addition, it took them more time to detect they had made an error, although they did not make significantly more errors.

Younger participants generally made shorter glances at the road than did older participants (0.9 s versus 1.6 s), but glances at the display were only slightly longer (1.4 s versus 1.1 s).

There were no significant age differences in most driving performance measures except for an increased number of lane departures by older participants (38.3% of addresses as opposed to 11.5% of younger drivers). In addition, all older participants committed at least one lane departure during the experiment whereas only 2 young men and 5 young women committed departures.

No significant differences were found in the difficulty and safety ratings given by younger and older participants, although there was a trend for older participants to rate tasks as more difficult and less safe.

Discussion of Experimental Limitations

Validity of the driving task

A serious effort was made to make the driving experience as realistic as possible within the limitations of the driving simulator. For example, whenever the vehicle departed the lane, the steering wheel vibrated to simulate driving on gravel and a corresponding sound was played. In addition, to negate the sense of security that one might have in a simulator, a penalty of $1 was given for each lane departure. The validity of relative results should be given more emphasis than the validity of absolute results.

A clear advantage of performing this experiment in the simulator was the ability to control workload levels. By creating roads of specific constant road curvature, and by controlling the relative position of traffic (up to 4 cars), the visual demand of driving was controlled, which could not have been done as accurately if the experiment were performed on the road. However, since driving workload levels were controlled, the experiment became more sterile than it would have been on the road and the workload levels were limited in range. Moreover, the range of workload levels examined was limited to those resulting from road curvature but did not include other characteristics such as unexpected traffic events or changes in the road scene that could elevate workload. It is the authors' view that in this case, the benefits outweighed the limitations.

Realism of the destination entry task

Although the entry methods relied on different modalities (visual and motor for the touch screen and speech for speech recognition), the feedback modality was the same for all entry methods explored in this experiment. This was done to allow for comparison between the methods without confounding the feedback modalities. Combinations of speech input and visual feedback are likely to be used in aftermarket systems in which speech recognition capabilities are added to existing visual displays. However, it is most likely that future speech recognition systems will rely on auditory feedback, either instead or in addition to the visual feedback. Thus, results and conclusions pertaining to the speech recognition entry methods are only applicable to systems that use visual, rather than auditory, feedback.

The speech recognition system used in this experiment was not an off-the-shelf system. However, since it was designed to simulate typical real-world systems, it is assumed to have been a good representation of such systems and the results in this regard are considered valid.

Although the simulated keyboard could be improved if it were enlarged and perhaps located more conveniently, its characteristics were chosen to resemble current in-vehicle systems. Another improvement commonly used in contemporary systems is dynamic elimination of characters from the keyboard based on what was previously typed and on what constitutes valid entries. While these interventions and others may improve the device by reducing task completion time, they are not likely to change the general findings in this report.

An additional difference between the in-vehicle task performed in this experiment and possible real-world tasks is that the participants read the address from a small display, representing a memo or an electronic address book. The position of the memo was fixed, while actual addresses are probably more varied in their presentation (from a piece of paper, written on the hand, on a cell phone, or memorized).

Sample of participants
The sample of participants was well controlled for age and gender but not for other factors such as typing or professional driving experience, intelligence, risk averseness, or other cognitive skills. However, the experimenters' experience and some of the data in this experiment show that the largest difference between individuals are either due to age or age-correlated characteristics, so age is the most appropriate and predominating individual factor. Nevertheless differences among participants were large enough to produce heterogeneous data. An additional shortcoming of using a small number of participants was that only 2 age groups were examined, though they were appropriate groups towards the extremes of licensed driver ages. Additional age groups may have allowed prediction of performance for a larger range of ages.

Although large performance differences were found between the 2 age groups, aging is not necessarily the only cause of these differences. Other factors, such as typing experience, driving experience, and risk averseness may have added to these differences.

Experimental setting
The experimental instructions required participants to enter addresses consecutively, while in real life they might not enter more than one address per trip. Furthermore, some of the participants would not have entered addresses while driving at all. Nevertheless, the repetitive nature of the data-entry tasks in this experiment compensated for lack of training, and provided ample data for analysis purposes.

Interpretation of results
Glance data analysis, error analysis, and key by key analysis of the results were performed for a subset of the participants due to the laborious nature of the reduction process. While such analyses had less statistical power, they provided important insight into the underlying processes.

Application and Implications

- Destination entry using speech recognition with visual feedback is much safer than with a keyboard but it is still not as safe (both objectively and subjectively) as only driving. While it certainly reduces the risk and its duration, this experiment proves that not all speech recognition systems should be considered risk free.

- Destination entry using a keyboard is unsafe and should be restricted. The combined probability of departing a lane on all road curvatures while entering a single address was about 20%. Additionally, it was rated as very difficult and extremely unsafe by most participants. It should be noted, however, that some of the participants (all young) had no problem performing the task without significant deterioration of their driving performance. However, since there are no restrictions on the age of vehicle buyers, vehicles should be designed for drivers of all ages.

- As the keyboard-entry task progressed, the lateral control of the vehicle deteriorated. There was no point, however, after which the rate of deterioration increased. It is therefore difficult to choose a task duration that discriminates between safe and unsafe driving.

- Typing errors and speech recognition errors have significant effects on task completion time and consequentially on driving performance. Task completion time in word dictation mode increased by 9 s (+71%) when a single recognition error was made. A similar effect was found in the other two modes tested (character spelling: 17 s, +46% and keyboard entry: 11 s, +20%). At the design level, special care should be made to reduce the probability of errors and to minimize the consequences of such errors.

- Since large performance differences due to driving workload and visual demand were noted, it is highly recommended that future experiments include driving workload (not necessarily road curvature) as an experimental factor or at least to keep it controlled.

- The Wizard of Oz technique proved very useful for the purposes of this experiment and is recommended for similar experiments in the future. However, close attention to details in constructing the simulated device is required. Most important is the realism of the simulated device. Reasonable and consistent system times as well as predetermined responses to unexpected events have to be established. In addition, every effort should be made to hide "the wizard" from the user (i.e., invisible and inaudible).

REFERENCES

British Standards Institution (1996). Guide to In-Vehicle Information Systems (Draft Document DD235:1996), London, U.K.: British Standards Institution.

Brooks, A., Lenneman, J., George-Maletta, K., Hunter, D.R., and Green, P. (1999). Preliminary Examinations of the Time to Read Electronic Maps: The Effects of Text and Graphic Characteristics, (Technical Report UMTRI-98-36, ITS RCE report #939418), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Campbell, J.L., Carney, C., and Kantowitz, B.H. (1998). Human Factors Design Guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO) (Technical Report FHWA-RD-98-057), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration.

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

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

Chiang, D.P., Brooks, A.M., and Weir, D.H. (2000). An Experimental Study of Destination Entry with an Example Automobile Navigation System (SAE Technical Paper No. 2001-01-0810). Warrendale, PA: Society of Automotive Engineers

Coleman, M.F., Loring, B.A., and Wiklund, M.E. (1991). User Performance on Typing Tasks Involving Reduced-Size, Touch Screen Keyboards, Vehicle Navigation and Information Systems Conference Proceedings (VNIS '91), New York: Institute of Electrical and Electronics Engineers, 534-549.

Goodman, M., Bents, F.D., Tijerina, L., Wierwille, W., Lerner, N., and Benel, D. (1997). An investigation of the safety implications of wireless communication in vehicles (Technical Report DOT HS 808 635). Washington, DC: U.S. Department of Transportation.

Green, P. (1999a). The 15-Second Rule for Driver Information Systems, ITS America Ninth Annual Meeting Conference Proceedings, Washington, D.C.: Intelligent Transportation Society of America, CD-ROM.

Green, P. (1999b). Navigation System Data Entry: Estimation of Task Times (Technical Report UMTRI-99-17), Ann Arbor, MI, The University of Michigan Transportation Research Institute.

Green, P. (1999c). Navigation System Data Entry: Estimation of Task Times (Technical Report UMTRI-99-17), Ann Arbor, MI, The University of Michigan Transportation Research Institute.

Green, P. (1999d). Visual and Task Demands of Driver Information Systems (Technical Report UMTRI-98-16), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P. (2000). The Human Interface for ITS Display and Control Systems: Developing International Standards to Promote Safety and Usability, invited paper presented at the International Workshop on ITS Human Interface in Japan, Utsu, Japan.

Green, P. (2001a). Driver Eye Fixations (chapter 5), in Dewar, R.E. and Olson, P.L. (eds.), Human Factors in Traffic Safety, Tucson, AZ: Lawyers and Judges Publishing (in preparation).

Green, P. (2001b). Synopsis of Driver Interface Standards and Guidelines for Telematics as of Mid-2001 (Technical Report UMTRI-2001-23), Ann Arbor, MI, The University of Michigan Transportation Research Institute.

Green, P. (2001c). Variations in Task Performance Between Younger and Older Drivers: UMTRI Research on Telematics, paper presented at the Association for the Advancement of Automotive Medicine Conference on Aging and Driving, February 19-20, 2001, Southfield, Michigan.

Green, P. and Wei-Haas, L. (1985). The Rapid Development of User Interfaces: Experience with the Wizard of Oz Method. Proceedings of the Human Factors Society-29th Annual Meeting, 470-474. (Accompanied by Green, P., Wei-Haas, L., Reifeis, S., and Ottens, D. A Brief Demonstration of the Wizard of Oz Rapid Prototyper (videotape).)

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

ISO/DIS 15007-1 (1997). Road Vehicles -- Measurement of Driver Visual Behavior with respect to transport information and control systems -- Part 1: Definitions and metrics, Geneva, Switzerland: International Standards Organization.

Japan Automobile Manufactures Association, (2000, February 22). Guideline for In-vehicle Display Systems - Version 2.1, Japan.

Malkin, F.J. and Christ, K.A. (1985). A Comparison of Voice and Keyboard Data Entry for a Helicopter Navigation Task. (HEL-TM-17-85), Aberdeen Proving Ground, MD: U.S. Army Research Laboratory.

Manes, D., Green, P., and Hunter, D. (1998). Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models, (Technical Report UMTRI-96-37, also released as EECS-ITS LAB FT97-077), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Nowakowski, C., Utsui, Y., and Green, P. (2000). Navigation System Evaluation: The Effects of Driver Workload and Input Devices on Destination Entry Time and Driving Performance and Their Implications to the SAE Recommended Practice (Technical Report UMTRI-2000-20), Ann Arbor, MI, The University of Michigan Transportation Research Institute.

Noy, Y.I (ed.) (1997). Ergonomics and Safety of Intelligent Driver Interfaces, Mahway, NJ: Lawrence Erlbaum Associates.

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

Olson, J.R. and Nilsen, E. (1987-1988). Analysis of the Cognition Involved in Spreadsheet Software Interaction, Human-Computer Interaction, 3, 309-349.

Parkes, A.M., and Franzen, S. (eds.) (1993). Driving Future Vehicles, London, UK: Taylor and Francis.

Richardson, B. and Green, P. (2000). Trends in North American Intelligent Transportation Systems: A Year 2000 Appraisal (Technical Report 2000-9), Ann Arbor, MI, The University of Michigan Transportation Research Institute.

Ross, T., Vaughan, G., Engert, A., Peters, H., Burnett, G., and May, A. (1995). Human Factors Design Guidelines for Information Presentation by Route Guidance and Navigation Systems (HARDIE Harmonisation of ATT Roadside and Driver Information in Europe Drive II Project V2008, Deliverable no. 19, Workpackage No L2)

Schwab, J.C. (1971). Method Time Measurement (Section S, Chapter 2), Maynard, H.B. (ed.) Industrial Engineering Handbook, New York: McGraw-Hill.

Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., and Ward, J.L. (1967). The Attentional Demand of Automobile Driving, Highway Research Record # 195, Washington, D.C.: National Academy of Sciences, Transportation Research Board, 15-33.

Society of Automotive Engineers (1999). Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques, (SAE recommended Practice J2396), Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers (2000). Navigation and Route Guidance Function Accessibility while Driving, (SAE Recommended Practice J2364), version of January 20, 2000, Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers (2001). Calculation of the Time to Complete In-Vehicle Navigation and Route Guidance Tasks (SAE Recommended Practice J2365), version of April 20, 2001, Warrendale, PA: Society of Automotive Engineers.

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, also released as EECS-ITS LAB FT97-077), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Tsimhoni, O. and Green, P. (1999). Visual Demand of Driving Curves Determined by Visual Occlusion, paper presented at the Vision in Vehicles 8 Conference, Boston, MA, August 22-25, 1999.

Tsimhoni, O., and Green, P. (2001). Visual Demand of Driving and the Execution of Display-Intensive In-Vehicle Tasks, Proceedings of the Human Factors and Ergonomics Society 45 Annual Meeting (to appear).

Tsimhoni, O., Yoo, H., and Green, P. (1999). Effects of Workload and Task Complexity on Driving and Task Performance for In-Vehicle Displays as Assessed by Visual Occlusion (Technical Report UMTRI-99-37), Ann Arbor, Michigan: The University of Michigan Transportation Research Institute.

Wakita, T. and Terashima, R. (1999). Visual Behavior Model for Navigation System Operation while Driving. Proceedings of the Sixth World Congress on Intelligent Transport Systems (CD-ROM), Washington, D.C., ITS America.

Wierwille, W.W., Antin, J.F., Dingus, T.A., and Hulse, M.C. (1988). Visual Attentional Demand of an In-Car Navigation Display System, in Gale, A.G., Freeman, M.H., Haslegrave, C.M., Smith, P., and Taylor, S.P., (eds.), Vision in Vehicles II, Amsterdam, Elsevier Science, 307-316.

Wooldridge, M., Bauer, K., Green, P., and Fitzpatrick, K. (2000). Comparison of Workload Values Obtained from Test Track, Simulator, and On-Road Experiments, paper presented at the Transportation Research Board Annual Meeting, Washington, D.C.

APPENDIX A. Summary of Prior UMTRI Research

Workload Assessment

Commonly, classes of workload measures are identified: (1) primary task performance (such as lane departures and standard deviation of lane position), (2) secondary task performance (such as response time to an in-vehicle prompt), (3) subjective ratings (such as the NASA Task Loading Index-TLX), and (4) physiological measurements (such as heart rate variability). Each of these classes of measures has its advantages and disadvantages, but none has the desired combination of high reliability, sensitivity, ease of measurement, and high correlation with crash experience.

Tsimhoni and Green (1999) -
Visual Demand of Driving Curves Determined by Visual Occlusion

Tsimhoni and Green (1999) (see also Wooldridge, Bauer, Green, and Fitzpatrick, 2000) utilized an approach based on the visual occlusion method to assess workload, an approach first proposed by Senders, Kristofferson, Levison, Dietrick, and Ward (1967). Senders observed that when driving on a straight road with little oncoming traffic in heavy rain, occlusion of all but the small sector cleared by the windshield wiper led to a "psychological speed limit." Up to that speed, there was no anxiety. Above that speed the driver became anxious and had to slow down.

In other words, in order to drive, drivers must see the road. Thus, in the simplest form of the visual occlusion method, drivers are told to close their eyes whenever they can while driving. The greater the fraction of time their eyes are open, the more difficult the driving task and the greater the perceived visual demand. In practice, measuring eye lid closure is not easy, so instead drivers are asked to press a button to get an automatically controlled 0.5 second glimpse of the road, a typical glance time. Button presses are easy to record. In prior studies, pressing a button caused the projected image (in a simulator) to change from gray to a real scene for 0.5 seconds.

Tsimhoni and Green (1999) found that there was a high correlation between the fraction of time drivers needed to see the road (visual demand, a surrogate for workload) and the inverse radius of curvature (Figure 32). In addition, they also determined that the visual demand for a curve began to increase from the steady state value prior to a curve at a distance of about 150 meters, peaked at or near the curve entry, and leveled within about 100 meters from the peak. As shown in Figure 33, for a fixed deflection angle (the change in heading angle of a vehicle), increasing the curve radius also increases the curve length and therefore the duration over which workload is elevated.

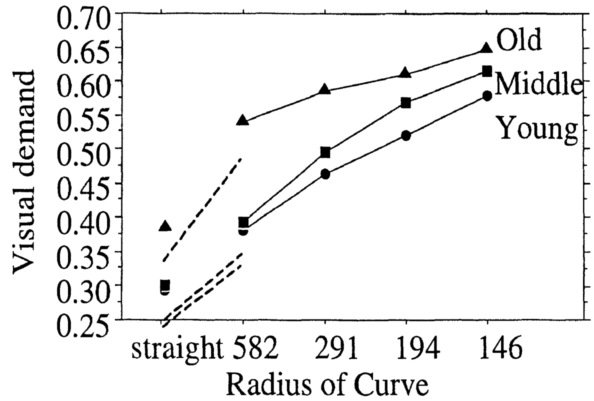


Figure 32. Mean visual demand as a function of curvature and age

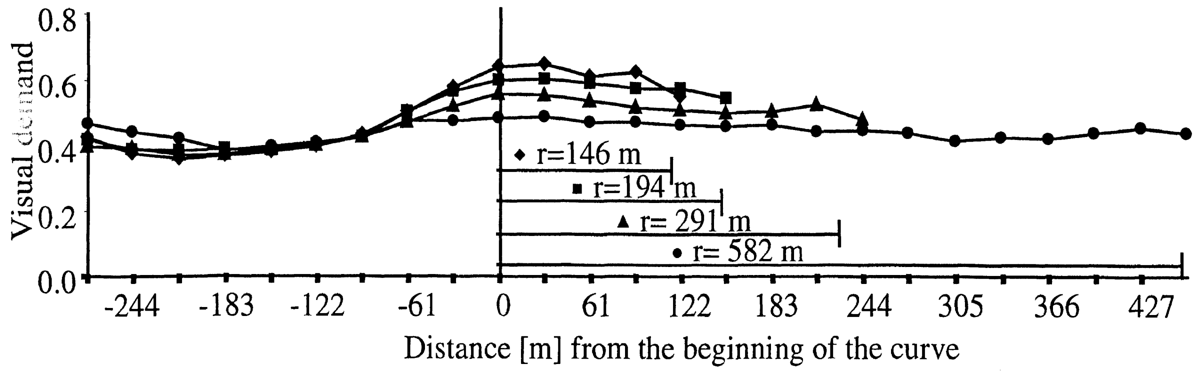


Figure 33. Visual demand for curves of different radii

Display-Intensive Tasks

Tsimhoni, Yoo, and Green (1999) - Effects of Workload and Task Complexity on Driving and Task Performance for In-Vehicle Displays as Assessed by Visual Occlusion

Utilizing this method, the authors conducted an experiment to explore how map reading and driving performance were simultaneously affected by workload (Tsimhoni, Yoo, and Green, 1999; Tsimhoni and Green, 2001). In that experiment, 16 participants (8 ages 21-28 and 8 ages 66-73) drove a simulator on roads with long curves of several different radii. Those radii were identical to those of Tsimhoni and Green (1999) except that the curves were longer.

Map-reading tasks were initiated 200 m (about 10 seconds) after the beginning of the curve, a point at which workload was fairly stable. All maps contained 12 streets labeled with common first names in the U.S. (Mark, Charles, Katie, etc.), and icons for a hotel, gas station, and fast food restaurant. Also, as is common on U.S. maps, a railroad and river were shown. Data supporting the selection of characteristics appear in Brooks, Lenneman, George-Maletta, Hunter, and Green (1999). Maps were shown on a center console display. The tasks included identifying (1) the street being driven, (2) the street on which the fast food restaurant (or one of two other categories) was located, and (3) the street intersecting a particular street on which a gas station (or one of two other categories) was located. In each case, participants responded by saying the street name. Map-reading tasks were completed both with the vehicle parked and while driving.

As in previous studies, as visual demand/workload increased (i.e., curve radius decreased), driving performance (as measured by the standard deviation of lane position and lane excursions) declined. Further, participants also made shorter and more glances to the display, but waited longer between glances. The net effect was that eyes-off-the-road time (total glance duration to inside the vehicle) decreased slightly as visual demand increased.

Overall, task completion time increased when the task was performed while driving (versus while parked), except for short single-glance tasks (or tasks completed under 3 seconds while parked), where task time decreased. This is believed to be due to time pressure to complete the task quickly. However, while driving, task completion times were relatively unaffected by the driving workload (e.g., low versus moderate workload).

Manual Input Tasks

Steinfeld, Manes, Green and Hunter (1996) -
Destination Entry and Retrieval with the Ali-Scout Navigation System

Destination entry has been the topic of several prior UMTRI studies. In Steinfeld, Manes, Green, and Hunter (1996), 36 drivers (12 ages 18-30, 12 40-55, 12 over 65, equally divided between men and women) entered and retrieved destinations into a Siemens Ali-Scout navigation system while seated in a vehicle mockup. Figure 34 shows the interface as it was configured in the experiment. The "found" key, added in the simulation, was used to indicate when the task was completed.

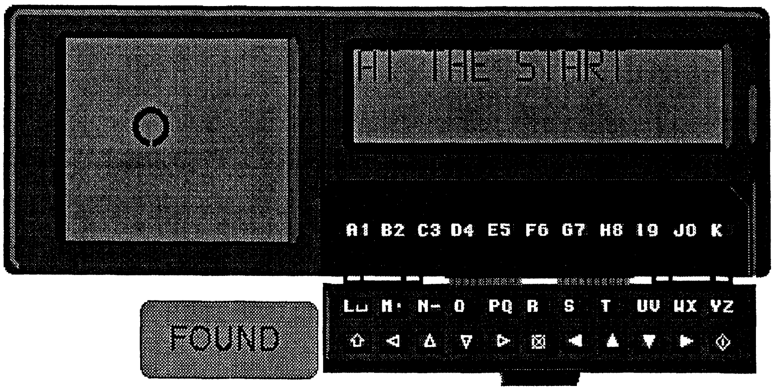


Figure 34. The simulated display unit with the door down

Each participant completed 30 trials, 20 with the real unit (10 under simulated dusk lighting conditions and 10 at night) and 10 with a touch-screen simulation of the Ali-Scout. Half of the 30 trials involved entry of destinations and half involved retrieval of destinations by scrolling through a list of saved destinations in the database. Address entry was unusual in that it involved input of the longitude and latitude of the address, not the state, city, street, and building number as typically is the case.

Figure 35 shows the distribution of entry and retrieval times for all participants and conditions. Notice that entry times for this interface were about 6 times longer than retrieval times. The total times reported are consistent with prior studies reviewed by Steinfeld, Manes, Green, and Hunter (1996).

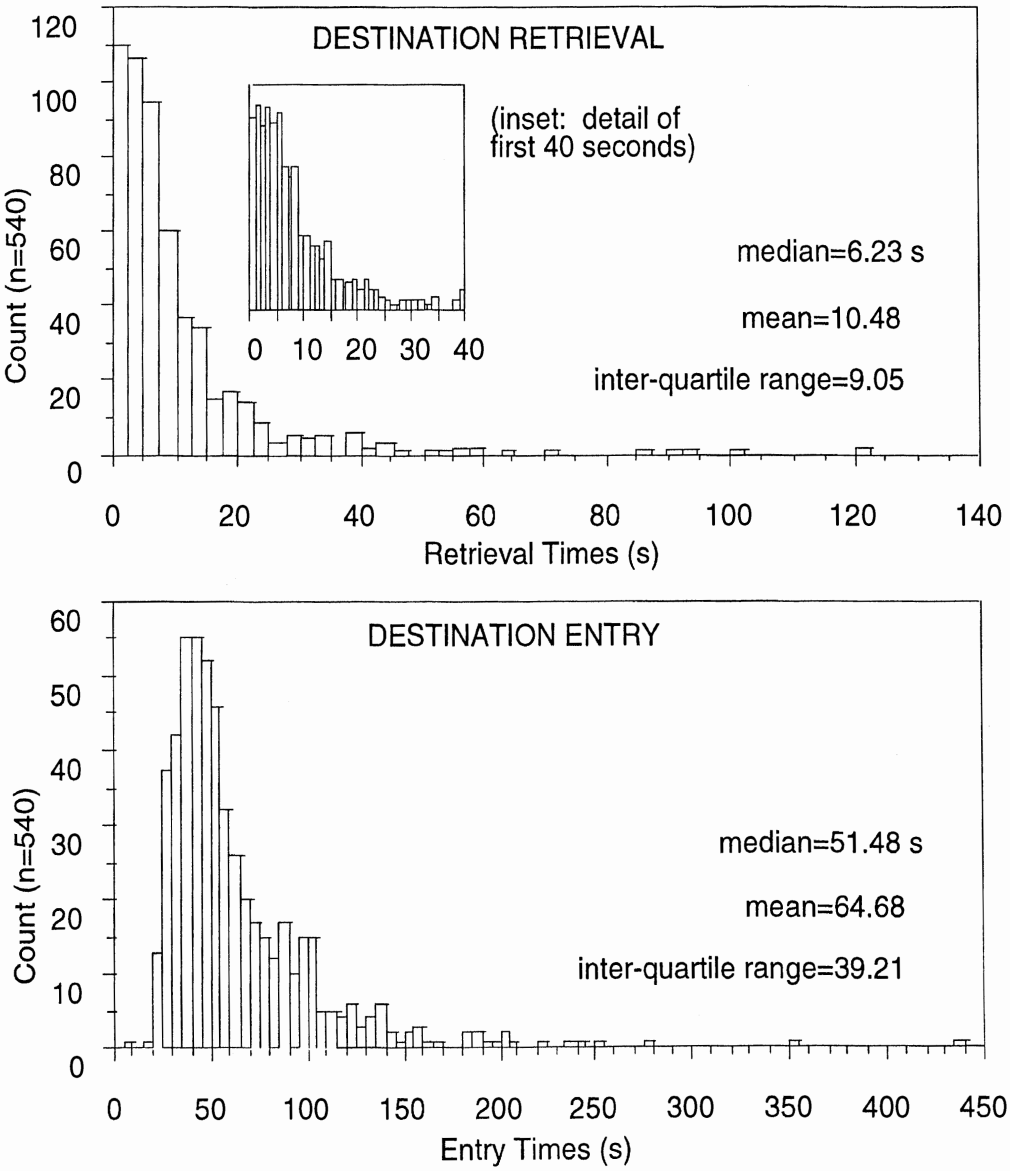


Figure 35. Distributions of entry and retrieval times

Figure 36 shows the effects of age and sex on overall performance. Notice the substantial increase in time (a factor of over two) due to age.

Figure 36. Age-sex interaction plots for destination retrieval and destination entry

Manes, Green, and Hunter (1998) -
Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models

Manes, Green, and Hunter (1998) carried out a further analysis of the entry and retrieval tasks. They determined the mean times for each type of keystroke (Figure 37) and for mental activities, as well as examining the effects of age and other factors. Notice in Figure 37 that the effects of age are fairly consistent across keystroke type.

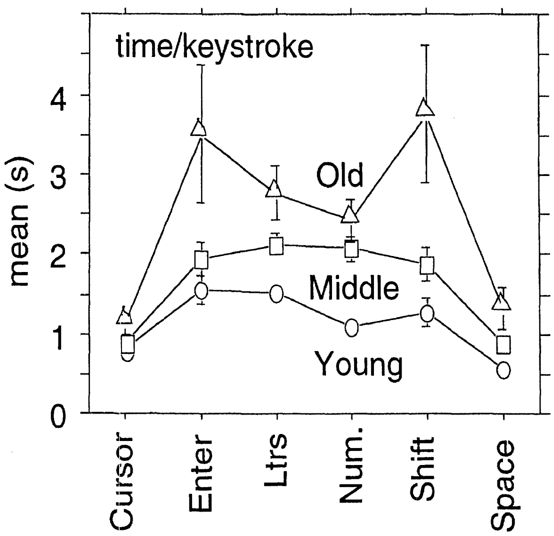


Figure 37. Time per keystroke reported by Manes, Green, and Hunter (1998)

Manes, Green, and Hunter (1998) also performed a detailed analysis of the time for keystroke as a function of the number of repetitions. This is important because some navigation interfaces require considerable use of cursor keys, and the time decreases when a key is pressed repeatedly (Table 18).

Using these data, Manes, Green, and Hunter developed a four-step method for estimating task completion times. Step one is to determine the keying actions in the input process, and for each action, determine the appropriate time. Table 18 shows their suggested times based on the Ali-Scout data. As a footnote, the Ali-Scout interface had very small keys with poor tactile and auditory feedback relative to other navigation interfaces.

Table 18. Mean keystroke times

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

Step two is to multiply the keystroke times to correct for the age of the user group (1.0 for young, 1.4 for middle aged, 2.2 for older drivers).

Step three is to correct for the lighting conditions. Multiply by 0.94 for dusk and 1.06 for night. (In many practical evaluations of task time, the effect of lighting is ignored.)

In step four the times for mental operations are inserted, 2.22 s per operation. Although not formally part of the process, Manes, Green, and Hunter note that estimates could be improved further by an overall linear correction. In general, the calculations tend to underestimate brief tasks (3 seconds or less) and overestimate very long tasks (over 60 seconds).

Nowakowski, Utsui, and Green (2000) -
Navigation System Evaluation: The Effects of Driver Workload and Input Devices on Destination Entry Time and Driving Performance and Their Implications to the SAE Recommended Practice

To examine the usability of a laptop navigation system, Nowakowski, Utsui, and Green (2000) had 16 drivers (8 ages 20-30, 8 ages 55-65) enter and retrieve destinations using a prototype navigation system. Figure 38 shows some of the entry screens. Both a keyboard and a hand-held remote control were used for entry.

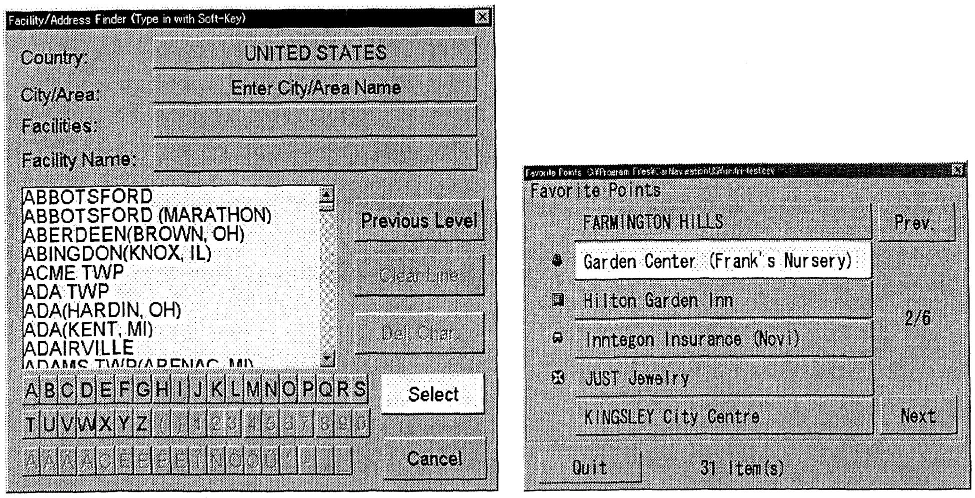


Figure 38. Example screens for address entry (left) and retrieval (right)

Table 19 shows the mean times for various tasks and task elements. Notice that the ratio of older to younger times is about 2.1 for keyboards, close to what the literature suggests (Green, 2001c; Manes, Green, and Hunter, 1998), but less for remote controls. The keying times for keyboard entry, available on the videotape only, were never reduced.

Table 19. Task mean times from Nowakowski, Utsui, and Green (2000)

Device	Keystroke/task name	Mean time		Older to younger
		Younger (s)	Older (s)	Ratio
Remote control	Cursor first keystroke	0.98	1.63	1.66
	Cursor additional keystroke	0.43	0.53	1.23
	Enter keystroke	0.99	1.54	1.55
	Overall mean keystroke times	0.80	1.23	1.53
	List selection total task time	21.70	32.50	1.50
Keyboard	List selection total task time	17.50	36.40	2.09
	Address entry total task time	70.80	145.80	2.06

Also examined was the impact of driving on task completion time. Notice, as shown in Figure 39, that driving did increase task times, but under the relatively low demands examined (3 and 6 degrees of curvature constant curves), the overall increase was only 2%. As expected, the correlation between static and dynamic performance was quite high (Figure 40). Interestingly, the primary effect seemed to be if one was driving or not, and not the particular level of workload.

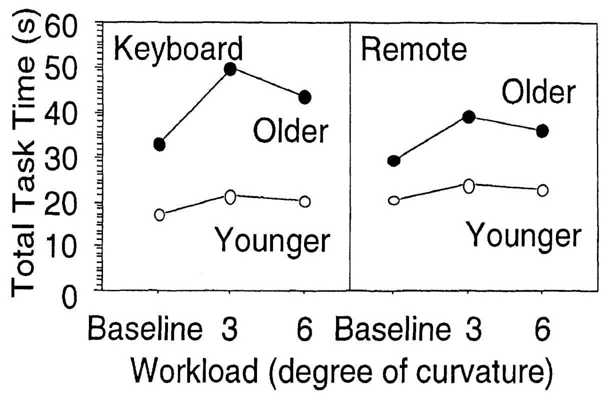


Figure 39. Relationship between task time and workload

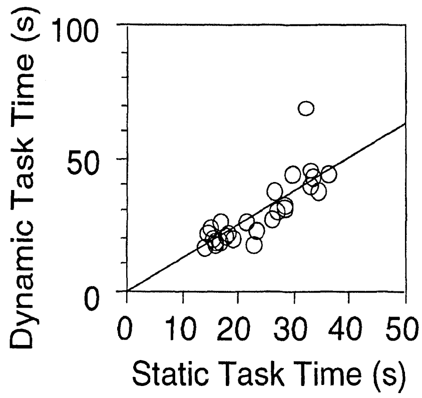


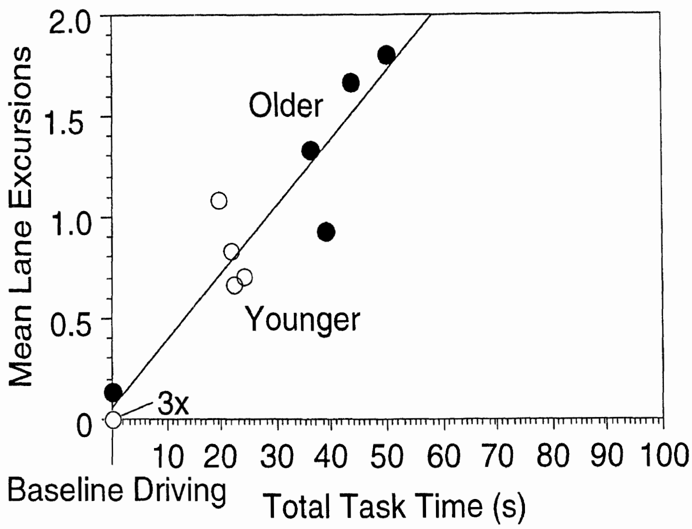
Figure 40. Relationship between static and dynamic task time

Finally, the static data were used to determine the times for the elemental actions when using the remote control. To facilitate analysis, the initial keystroke in each sequence was discarded. Remaining were 6,284 cursor keystrokes and 1,619 keystrokes from 14 of the 16 participants from whom usable data were obtained. On average, the times bracketed those in SAE J2365, the Society of Automotive Engineers (2001) Recommended Practice for estimating task times. Because some are high and some are low, they are expected to average out when a total task time is computed.

Table 20. Summary of keystroke times and comparison to SAE J2365 estimates

	Experiment findings (remote)		J2365 Estimate (keystrokes)	
	Younger (s)	Older (s)	Younger (s)	Older (s)
Cursor once	0.98 (+23%)	1.63 (+20%)	0.80	1.36
Cursor additional	0.43 (+8%)	0.68 (-6%)	0.40	0.72
Enter key	0.99 (-17%)	1.54 (-25%)	1.20	2.04

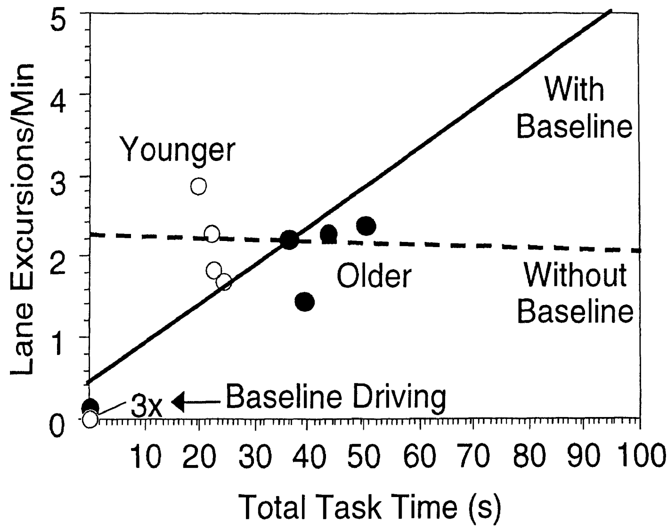
One of the more interesting findings in Nowakowski, Utsui, and Green (2000) was the relationship between driving performance and task time (Figures 10 and 11). Figure 41 shows the correlation averaging across the number of participants within each age group and the number of trials for each of the 6 tasks (baseline driving, remote control use, and keyboard use by 2 levels of driving demand). The longer a task takes, the greater the number of lane excursions, with the slope being specific to the road geometry tested.



MLE = .051 + .033 * TTT R^2 = .908

Figure 41. Mean lane excursions per trial as a function of total task time

In Figure 42, the data have been examined in terms of excursions per minute. Nowakowski, Utsui, and Green note that using lane excursions per minute might be a better driving performance measure because it eliminates the overrepresentation of lane excursions due to longer task completion times. The particular value determined for the effect of task duration on the rate of lane excursions depends upon whether the data points of baseline driving are included. If included, lane excursions increase by 0.05 excursions per minute. If only the excursions that occurred when performing the in-vehicle task are included (dotted line), the rate of excursions is independent of total task time. As a rough approximation, in both cases, the rate is fairly constant, about 2 excursions per minute for the roads examined.



With Baseline: LE/Min = 0.4 + 0.05 * TTT $R^2 = 0.642$

Without Baseline: LE/Min = 2.26 - 0.01 * TTT $R^2 = 0.03$

Figure 42. Lane excursions per minute as a function of total task time

Society of Automotive Engineers (2001) -
Calculation of the Time to Complete In-Vehicle Navigation and Route Guidance Tasks

One result of these studies on destination entry has been the development of SAE Recommended Practice J2365, a method for estimating navigation system task times. Since this project began, J2365 has gone from a proposal to a document approved by the SAE Safety and Human Factors Committee. The times for elemental actions in that document (Table 21) were derived from four primary sources: (1) the classical Keystroke-Level Model Operators described by Card, Moran, and Newell (1980, 1983) obtained from a variety of office tasks, (2) data from Olson and Nilsen (1997-1998) for spreadsheet use, (3) data from Manes, Green, and Hunter (1998) for entering data into a Siemens Ali-Scout Navigation System, and (4) reach times from Methods Time Measurement-1 (MTM-1), the most popular predetermined time system used by industrial engineers (Schwab, 1971).

Table 21. Operator times (seconds) from SAE J2365

Code	Name	Operator description	Time (s)	
			Younger (18-30)	Older (55-60)
Rn	Reach near	from steering wheel to other parts of the wheel, stalks, or pods	0.31	0.53
Rf	Reach far	from steering wheel to center console	0.45	0.77
C1	Cursor once	press a cursor key once	0.80	1.36
C2	Cursor 2 times or more	time/keystroke for the second and each successive cursor keystroke	0.40	0.68
L1	Letter or space 1	press a letter or space key once	1.00	1.70
L2	Letter or space 2 times or more	time/keystroke for the second and each successive cursor keystroke	0.50	0.85
N1	Number once	press the letter or space key once	0.90	1.53
N2	Number 2 times or more	time/keystroke for the second and each successive number key	0.45	0.77
E	Enter	press the Enter key	1.20	2.04
F	Function keys or shift	press the function keys or shift	1.20	2.04
M	Mental	time/mental operation	1.50	2.55
S	Search	search for something on the display	2.30	3.91
Rs	Response time of system-scroll	time to scroll one line	0.00	
Rm	Response time of system-new menu	time for new menu to be painted	0.50	

Note 1: The keystroke times do not include the time to move between keys and are thus different from the keystroke times shown in Table 18 above.
Note 2: System response times to show new menus may be determined empirically.

Thus, the literature to date provides initial estimates of the mean times for various manual operations, and shows a strong relationship between task times when parked and while driving.

APPENDIX B. Error Analysis and Predicting Task Completion Time

Detailed analysis of 4 participants, 1 from each age-gender category

Speech Recognition in Word Dictation Mode

When using speech recognition in word dictation mode, participants said the address word by word and verified that words were registered correctly by the system as they appeared on the in-vehicle display. If a word was unrecognized, they said "scratch that" to delete the word and correct it. After the address was completed, the system indicated that the address was accepted (Figure 43).

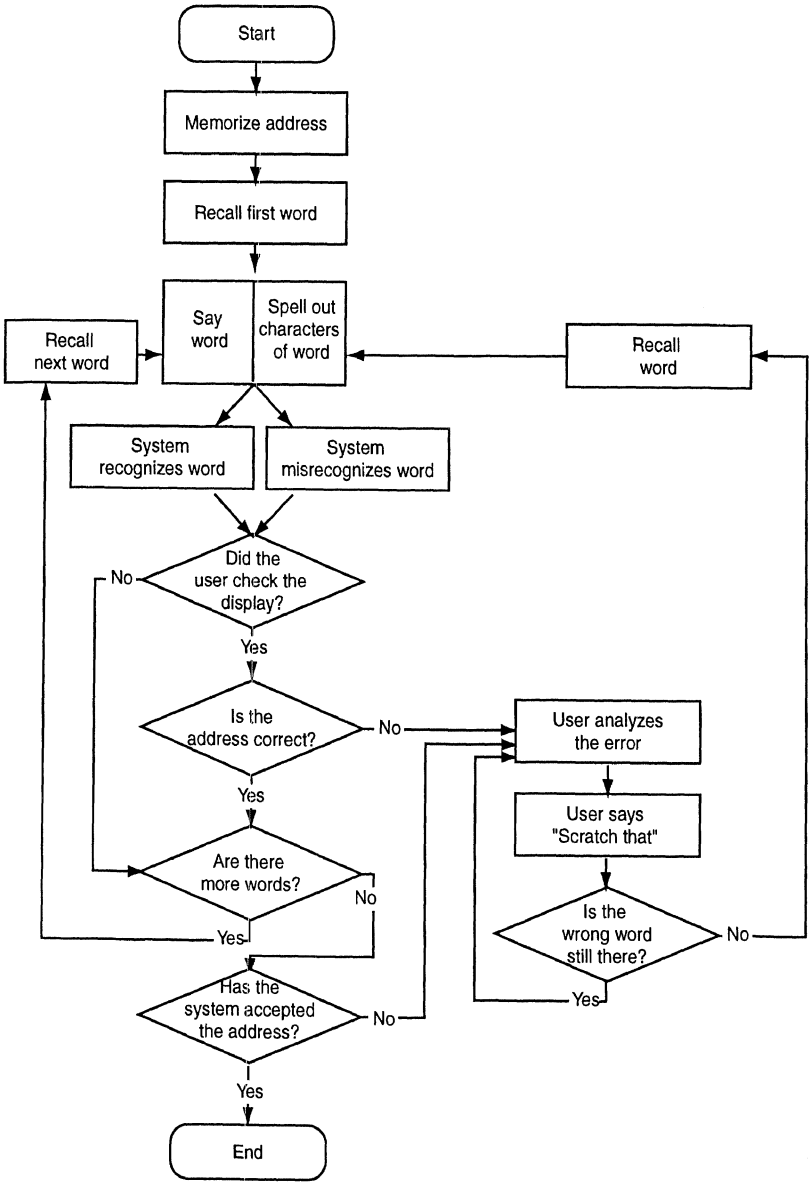


Figure 43. Flow chart of the process of data entry using recognition

For each participant, the mean time to enter a single word was calculated from 48 words in 12 addresses. The mean time to detect an error, say “scratch that”, and correct it (error correction time) was measured. Thus, total task completion time (TCT_{WD}) as shown in Equation 5 was based on the time per word, the number of errors, and error correction time. (The time per word included the time to say the word and the system recognition time.)

$$TCT_{WD(s)} = \{ \sum_{words} \text{Time per word}_{(s)} \} + \sum_{Errors} (\text{Error correction time}_{(s)} + \text{Time per word}_{(s)}) \qquad (5)$$

An alternative view of the data (Equation 6) considers multiple errors. If errors are independent events, the probability of 2 errors for the same word is the square of the probability of 1 error. At reasonable error rates (10% per word), the probability for 2 errors to occur for the same word is very low (1%). However, at higher error rates (25%) it cannot be ignored (6%). (The probability for 3 or more errors to occur is so low even for a base error rate of 25% (<1.6%), that this case was ignored.)

$$\text{Number of errors} = (P_{error} + P_{error}^2) \cdot \text{Number of words} \qquad (6)$$

Table 22 presents the mean time per word and mean correction time (to detect and correct a single error) as well as the predicted task completion time for 4 participants, one from each age and gender group. Since no significant difference was found between the parked condition and driving on a sharp curve, data for both conditions were combined.

Table 22. Empiric and predicted values for word entry

		Age - gender group (1 from each)				Mean across 4 participants
		Young female	Young male	Old female	Old male	
Empiric mean values	Mean time per word [s]	2.6±0.5	3.3±1.0	2.8±0.7	4.3±1.3	3.3±0.7
	Mean correction time (detect and correct 1 error [s])	6.0±0.7	7.5±1.3	5.8±0.2	7.7±1.3	6.8±1.0
Predicted from mean using Eq. 5	Mean entry time with no errors [s]	10.5	13.1	11.4	17.2	13.1
	Mean entry time with 1 error [s]	19.1	23.9	20.0	29.2	22.4

The predicted task completion times (Equation 5) were close to the actual times. For the no-error case, the prediction ranged from 98 to 102% (mean 100%) of the actual. For the one-error case, there was an over-prediction ranging from 106 to 114% (mean 110%). Figure 44 shows the differences between the predicted and measured values for the no-error case and for the one-error case as well as the predicted values for 10% error per word.

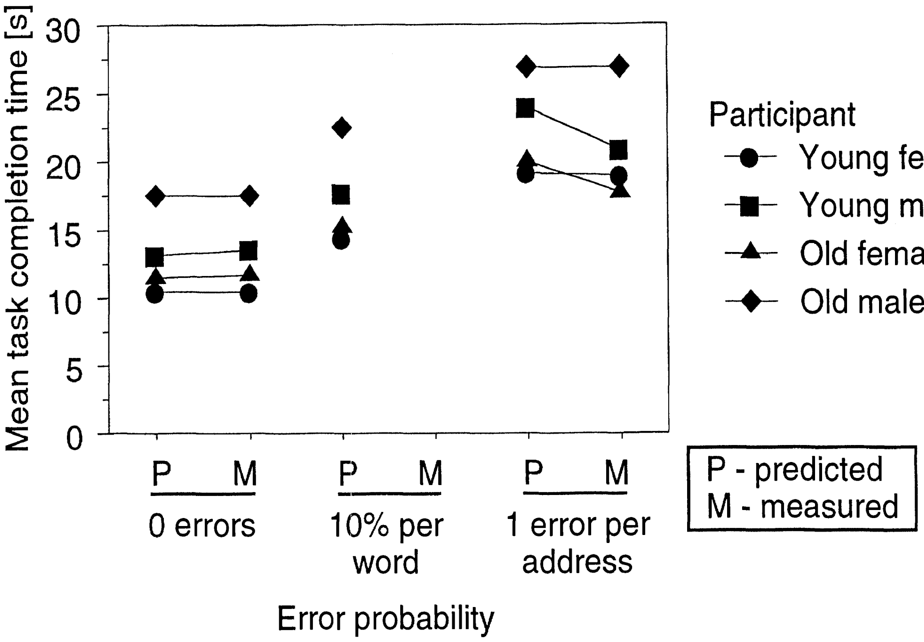


Figure 44. The effect of number of errors on task completion time

Speech Recognition in Character Spelling Mode

When using speech recognition in character spelling mode, participants entered the address character-by-character and verified that words were displayed correctly. If a word was unrecognized, they said "scratch that" to delete the entire word and then proceeded to re-enter the address. As in word dictation mode, when all four words of an address had been entered correctly, the system indicated that the address was accepted. (See Figure 43 above for a flow chart of the process.)

For each participant, the mean time to enter a single character (time per character) was calculated from 72 words in 18 addresses (360 required characters). The number of errors and the mean time to detect an error and delete it by saying "scratch that" (error deletion time) were measured. The time for word recognition (system time) was 2 seconds in all cases. Thus, the total task completion time (TCT_{SP}), as shown in Equation 7, was based on the time per character, system time, and error deletion time.

$$TCT_{SP(s)} = \sum_{words} \{ \sum_{characters} \text{Time per character}_{(s)} + \text{System time}_{(s)} \} + \sum_{Errors} \{ \text{Error deletion time}_{(s)} + \sum_{character} \text{Time per character}_{(s)} + \text{System time}_{(s)} \} \quad (7)$$

As an example, the task completion time for character entry while driving for a young female (Table 24), who had an entry time per characters of 1.2 s and error deletion time of 4.7 s, is shown in Equation 8.

$$TCT_{SP(s)} = 4 \text{ words} * ((5 \text{ chars} * 1.2 \text{ s/char}) + 2.0 \text{ s}) + 0.25 \text{ errors/word} * 4 \text{ words} * (4.7 \text{ s} + 5 * 1.2 \text{ s} + 2.0 \text{ s}) = 44.6 \text{ s} \quad (8)$$

Since task completion time in the driving condition was longer than while parked (+11%), the data for each condition were analyzed separately. The mean character entry time, error deletion time, and predicted task completion time are presented in Table 23 for the parked condition and in Table 24 for the driving condition. The mean time per character while driving was about 10% longer than while parked. Similarly, the mean error deletion time increased, but the increase was not consistent across all participants.

Table 23. Empiric and predicted values for spelling entry—Parked

		Age - gender group (1 from each)				Mean across 4 participants
		Young female	Young male	Old female	Old male	
Empiric mean values	Mean time per character [s]	1.1±0.3	1.2±0.6	1.0±0.2	1.9±0.6	1.3±0.4
	Mean error deletion time [s]	8.6±7.1	3.8±0.2	3.3±0.2	3.6±0.4	4.8±2.5
Predicted from mean using Eq. 6 and Eq. 7	Mean entry time with no errors [s]	29.4	32.8	28.2	45.2	33.9
	Mean entry time with 1 error [s]	45.8	45.4	39.1	61.1	48.6

Table 24. Empiric and predicted values for spelling entry—Driving

		Age - gender group (1 from each)				Mean across 4 participants
		Young female	Young male	Old female	Old male	
Empiric mean values	Mean time per character [s]	1.2±0.5	1.3±0.6	1.2±0.5	2.1±1.0	1.5±0.5
	Mean error deletion time [s]	4.7±1.7	3.9±1.4	3.9±0.7	15.5±18.8	7.0±5.7
Predicted from mean using Eq. 6 and Eq. 7	Mean entry time with no errors [s]	31.4	34.0	31.8	50.4	36.9
	Mean entry time with 1 error [s]	44.6	47.1	44.3	79.6	53.9

The predicted task completion times (Equation 7) were close to the actual times. In the parked condition, the prediction ranged from 96 to 102% (mean 99%) and in the driving condition it ranged from 92 to 104% (mean 98%). Figure 45 shows the differences between the predicted and measured values while driving for the no-error and one-error case, as well as the predicted values for 10% error per word.

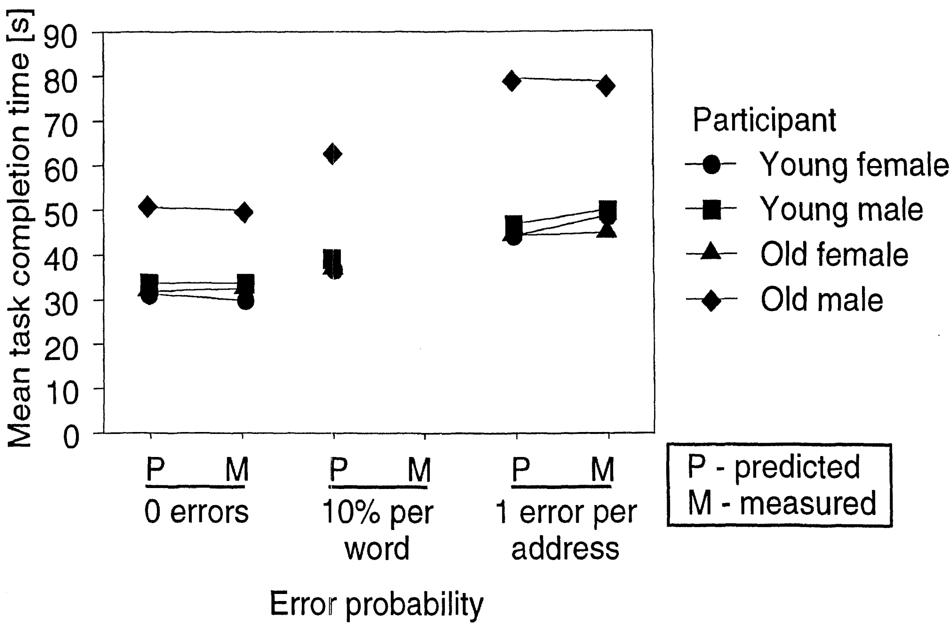


Figure 45. The effect of errors on task completion time in character spelling mode while driving

Keyboard Entry

Unlike speech recognition, in the keyboard-entry method the experimenter had no control over the errors that participants committed. Errors were categorized into 3 groups: double clicks, keys adjacent to the correct key, and any other key. For the 4 participants whose data were analyzed (1660 required key presses), the overall probability of committing an error was 4.7% per key, which amounted to about one error per address. The most common error was the double click (2.6%), followed by an adjacent key (1.3%) and other keys (0.8%). On average, participants detected an error within 3.5 s of making it (range: 1 to 16 s).

For each participant, the mean time to type a single character (time per key) was calculated. Correction time was predicted by counting the number of keys that had to be deleted after the error was noticed and the time it would take to re-type them. The total task completion time (TCT_{KB}) was then predicted using Equation 9. The typical parameter values are reported in Table 25.

$$TCT_{KB(s)} = \sum_{\text{Characters}} \text{Time per key}_{(s)} + \sum_{\text{Error types}} \text{Error probability} \cdot (\text{Detection}_{(s)} + \text{Correction}_{(s)}) \qquad (9)$$

Table 25. Typical parameter values (mean across participants) for keyboard entry

	Parked	Straight	Moderate	Sharp	P value
Time per key [s]	1.16±0.31	2.10±1.15	2.28±1.20	2.45±1.09	0.01
Error probability per key [%]	3.8±2.8%	5.0±3.9%	3.2±2.4%	7.2±5.3%	ns
Detection time [s]	3.7±3.8	3.0±3.0	3.0±4.7	4.4±4.3	ns
Keys to correct per error	3.6±2.5	2.2±1.3	3.3±2.5	2.2±1.0	ns
Correction time [s]	3.9±2.1	1.5±1.2	2.1±1.5	3.0±2.5	ns
TCT [s]	33±11	49±27	58±41	70±46	0.12

The predicted task completion times (Equation 8) were within 10% of the actual times (parked 105%, straight 96%, moderate curve 99%, and sharp curve 99%). Figure 46 shows the differences between the predicted and measured values for four participants under different driving workload conditions.

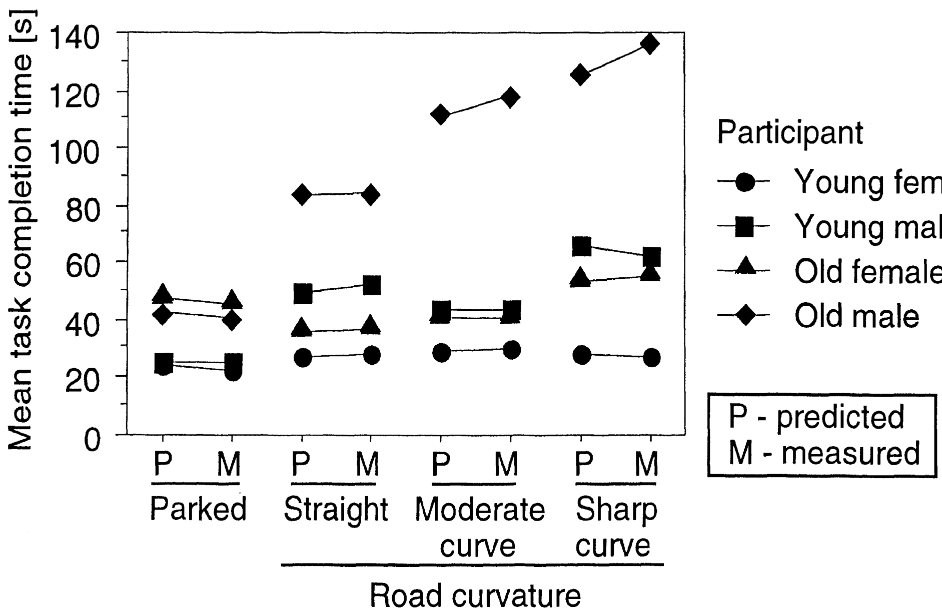


Figure 46. The effect of driving conditions on keyboard-entry task completion time

At least 3 possible factors could have attributed to the increase in task completion time as the driving workload increased: time between keys, higher error probability, or more time to detect the errors. Examination of the data revealed that while the time between keys increased more than twofold as workload increased, the error probability and the time to detect errors remained unchanged. Thus, the time between keys was the main reason for the increase in task completion time as the driving workload increased.

Figure 47 shows the distribution of the number of keys that were corrected for each error. More than 50% of the errors were detected immediately and therefore only one key was required to correct them. Errors that were detected after 5 keys or less accounted for 90% of all errors for these participants.

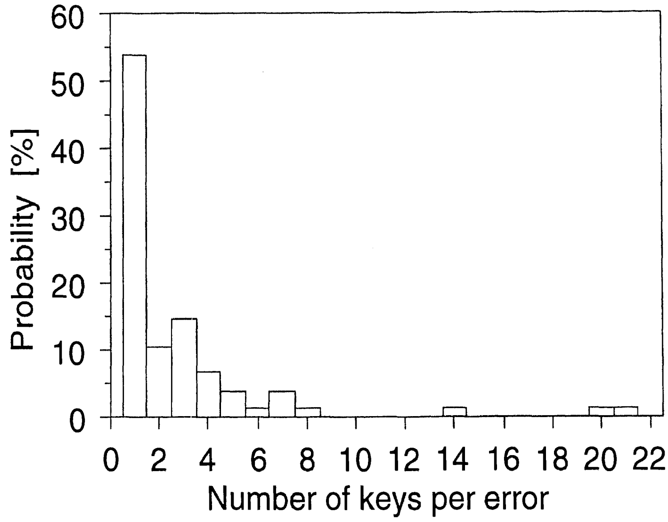


Figure 47. Distribution of the number of keys corrected per error

Touch-Screen Keyboard Error Analysis

To simplify the error analysis of all participants, the number of times the backspace key was pressed was used to characterize typing errors. (The number of backspaces is a function of the number of errors and the time to detect each of them.) On average, the backspace key was pressed 4 times per address (4.8 for older participants and 3.4 for younger participants). The effect of curvature on the number of backspaces was not statistically significant. The mean number of backspaces was 3.6 while parked and 4.1 on sharp curves.

The mean time spent on backspaces, however, was significantly affected by road curvature ($p=0.01$, Figure 48). Participants did not make more backspace keystrokes but they took more time between them. The time spent between consecutive backspaces was 2.1 s while parked and 4.1 s on sharp curves. Accordingly, the total time per address spent on error correction was 7.3 s while parked and 20.7 s on sharp curves, a difference of almost a factor of 3.

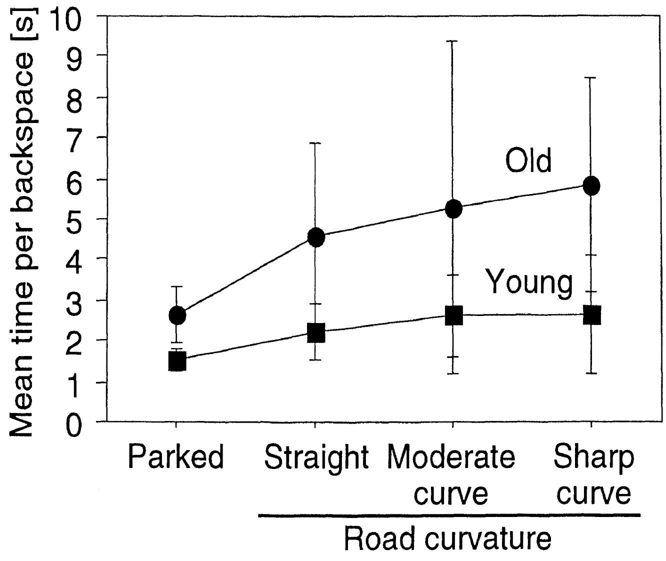


Figure 48. The mean time per error correction by road curvature

Predicting Task Completion Time

Reducing the data into simple equations allows prediction of task completion time with parameter levels that were not tested. For example, the effect of error rate on task completion time on a straight road in each of the 3 entry methods can be predicted using Equations 5, 7, and 9.

$$TCT_{WD(s)} = \{ \sum_{words} \text{Time per word}_{(s)} \} + \sum_{Errors} (\text{Error correction time}_{(s)} + \text{Time per word}_{(s)})$$

$$TCT_{wd} = 4 \text{ words} * 3.3 \text{ s} + (6.8 \text{ s} + 3.3 \text{ s}) * \# \text{ errors}$$
$$= 13.2 \text{ s} + 10.1 \text{ s} * \# \text{ errors}$$
$$= 13.2 \text{ s} + 10.1 \text{ s} * 4 (P_{error} + P_{error}^2) \qquad P_{error} = \text{Error probability per word}$$

$$TCT_{SP(s)} = \sum_{words} \{ \sum_{characters} \text{Time per character}_{(s)} + \text{System time}_{(s)} \} + $$
$$+ \sum_{Errors} \{ \text{Error deletion time}_{(s)} + \sum_{character} \text{Time per character}_{(s)} + \text{System time}_{(s)} \}$$

$$TCT_{sp} = (20 \text{ characters} * 1.5 \text{ s} + 4 \text{ words} * 2.0 \text{ s}) + (7.0 \text{ s} + 7.0 \text{ s} + 2 \text{ s}) * \# \text{ errors}$$
$$= 38 \text{ s} + 16 \text{ s} * \# \text{ errors}$$
$$= 38 \text{ s} + 16 \text{ s} * 4 (P_{error} + P_{error}^2) \qquad P_{error} = \text{Error probability per word}$$

$$TCT_{KB(s)} = \sum_{Characters} \text{Time per key}_{(s)} + \sum_{Error\ types} \text{Error probability} \cdot (\text{Detection}_{(s)} + \text{Correction}_{(s)})$$

$$TCT_{kb} = 24 \text{ characters} * 2.1 \text{ s} + \# \text{ errors} * (3.0 \text{ s} + 1.5 \text{ s})$$
$$= 50.4 \text{ s} + 4.5 \text{ s} * 24 * (P_{error} + P_{error}^2) \qquad P_{error} = \text{Error probability per key}$$

To predict task time as a function of recognition accuracy, correction times were desired. Using proper controls, prediction of error costs for any recognition accuracy was possible, not just for the rates explored. This should allow estimation of the overall benefits, in terms of task completion time, of voice recognition systems' accuracy. Figure 49 displays the relation between speech-recognition accuracy and task-completion time using Equations 7 and 9 based on data from 4 participants. The recognition rate in this experiment (92%) is highlighted with gray arrows indicating the resulting task completion times.

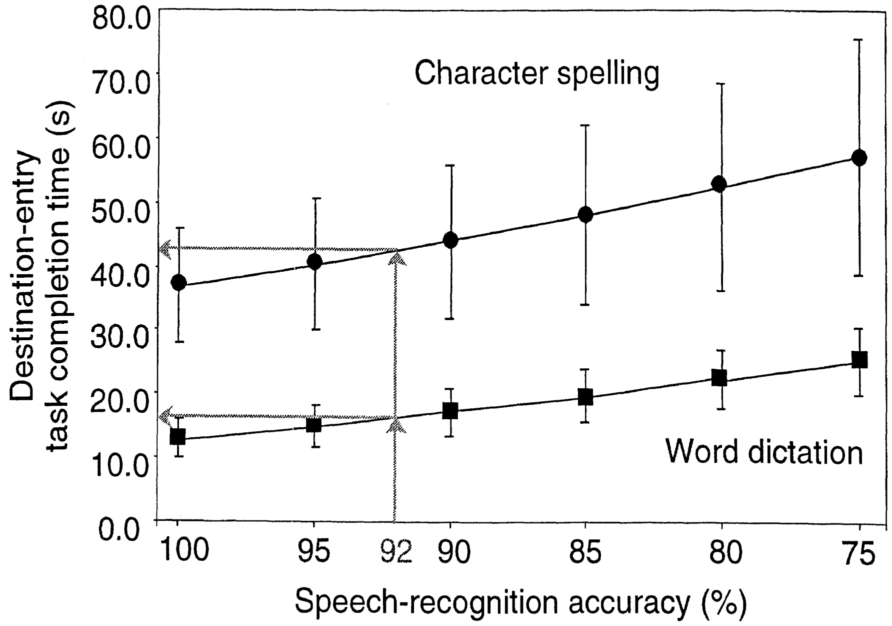


Figure 49. Task completion time as a function of speech-recognition accuracy

APPENDIX C. Time-per-Key

Time-per-key values for the prediction equation were aggregated across all types of keys. Table 26 and Figure 50 show the times per key as a function of curvature. The Enter and space keys were separated from the other keys. The difference between keys was marginally significant (p=0.05).

Overall, the ratio between times for older and younger participants was 1.78 (range 1.08 to 2.15). The ratio between driving and parked was 1.97 (range 0.88 to 2.56).

Table 26. Mean time per key values (enter key, space key, and other keys)

	Parked			Straight			Moderate curve			Sharp curve		
Age	Y	O	All	Y	O	All	Y	O	All	Y	O	all
Enter key	1.3	1.4	1.3	1.8	3.5	2.6	1.9	3.3	2.6	1.7	3.1	2.4
Space key	0.9	1.3	1.1	1.4	2.5	2.0	1.4	2.9	2.2	2.0	3.4	2.7
Other keys	0.8	1.3	1.1	1.3	2.1	1.7	1.3	2.8	2.1	1.4	3.2	2.3

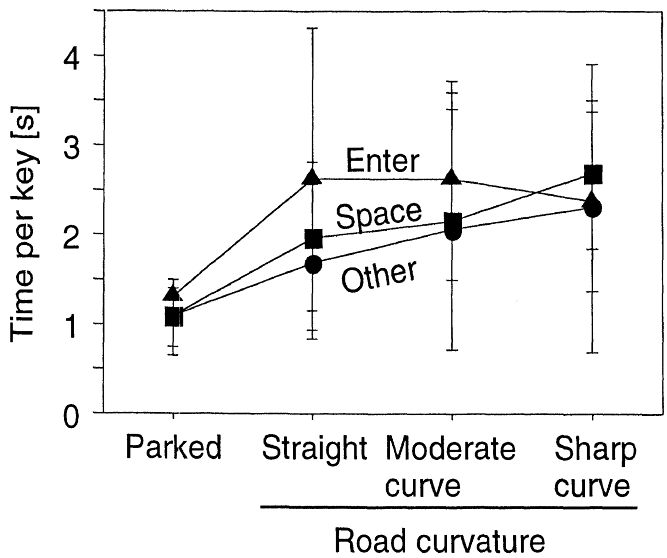


Figure 50. Time per key (enter key, space key, and other keys)

Table 27 shows the comparison of keystrokes from the current experiment with the SAE J2365 keystroke level estimation. Since the ages of the older participants tested were different from the 55-60 range suggested by J2365, the comparison contains an age correction. Based on a first order approximation of linearity with age, the age correction factor was changed by 0.0215 per year for 10 years from 1.7 to 1.915. The results in the current experiment are within a 10% range of the corrected J2365 estimations.

Table 27. Comparison of keystrokes and completion times to SAE J2365

	J2365 Estimate (Keystrokes)		Current experiment (4 Ss)	
	Younger (s)	Older (s) corrected for 65-70 (1.91x)	Younger (s)	Older (s)
Enter key (including reach far)	1.6	3.2	1.8 (+8%)	3.3 (+4%)
Letter (including reach far)	1.4	2.8	1.6 (+9%)	3.0 (+6%)
Space (including reach far)	1.4	2.8	1.3 (-8%)	2.7 (-3%)
Task completion time[1]	24.2	42.1	23.0 (-5%)	46.0 (+10%)

Note 1: Task completion times were based on all 24 participants.

The estimation of task completion time using J2365 was based on 16 characters, 4 digits, a space, and 3 Enter keys. (16*1.0 s+4*0.9 s+1.0 s+3*1.2 s=24.2 s)

APPENDIX D. Glance Behavior and Task Partitioning

Glance and Task Partitioning Measures

Glance behavior and character-entry behavior were analyzed as a function of road curvature and age. Table 28 contains the mean values with a range of one standard deviation. Glance behavior data was extracted from video analysis.

Table 28. Mean and standard deviation of glance and character entry behavior

		Road Curvature		Age	
		Straight	Sharp	Young	Old
	Task completion time [s]	50.0±25.4	83.2±44.9	41.4±15.6	91.8±40.8
Glance	Total glance duration[s]	30.1±10.9	34.0±13.3	26.1±10.6	38.0±10.7
Glance	Mean glance duration [s]	1.4±0.3	1.1±0.2	1.4±0.3	1.1±0.3
Glance	Time between glances [s]	1.0±0.5	1.4±0.6	0.9±0.3	1.6±0.5
Glance	Number of glances	21.9±8.6	33.1±14.7	20.6±10.7	34.5±11.9
Task Partitioning	Time between fields [s]	3.5±2.2	5.5±3.4	2.8±2.4	6.2±2.5
Task Partitioning	Time between groups [s]	2.6±1.0	3.3±1.4	2.1±0.7	3.7±1.2
Task Partitioning	Time within groups [s]	0.8±0.2	0.6±0.2	0.6±0.1	0.8±0.2

Distribution of Task Partitioning Measures

Figure 51 displays how the distribution of task partitioning measures changed from the straight road to the sharp curve. The distributions were grouped together for comparison among task partitioning measures. This makes it easy to see how much longer time between fields is than time within groups, for example. Time between fields and time between groups is based on the first 2 addresses per road curvature for the 12 participants, while time within groups is based on the mean of the first 2 addresses per road curvature for the 12 participants.

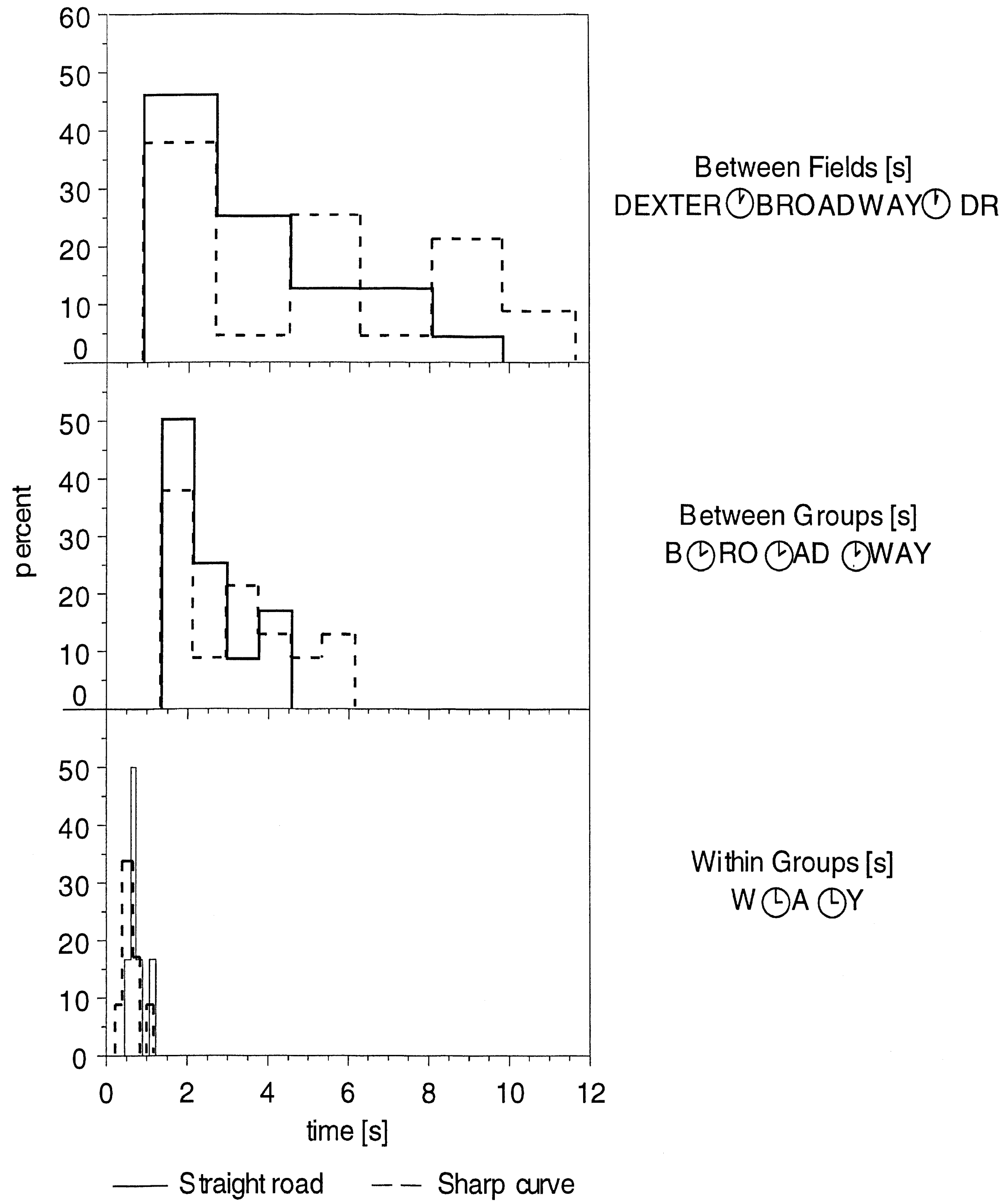


Figure 51. Distribution of task-partitioning behavior

Keyboard-Entry Method Keys per Glance Distribution

Figures 61 and 62 contain the distribution of the mean number of characters entered per glance at the display on a straight road and sharp curve, respectively. The number of keys pressed per glance were derived from the first two addresses of a straight road and sharp curve for the same 12 participants as used for all glance behavior and task partitioning analyses.

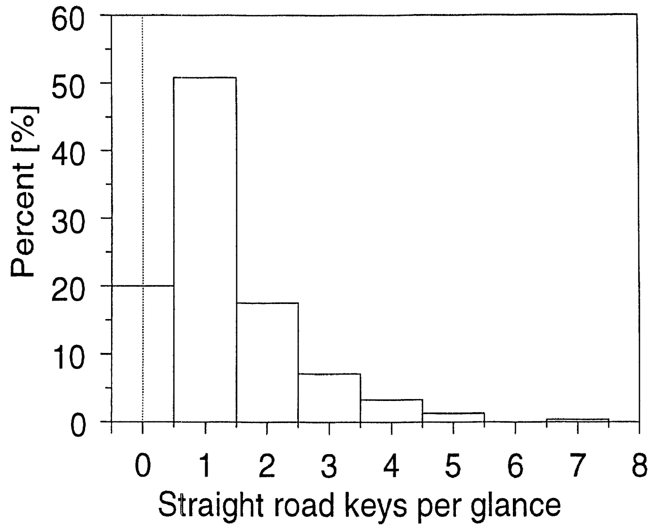


Figure 52. Entered keys per glance on straight road

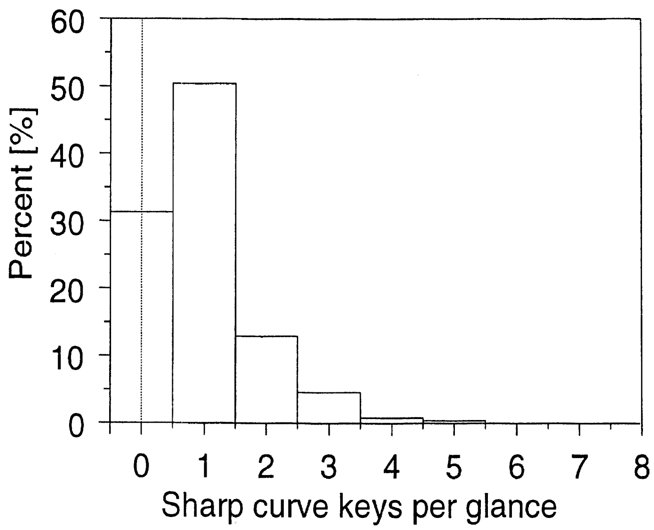


Figure 53. Entered keys per glance on sharp curve

APPENDIX E. Consent Form

Date: ____________ Participant number: ____________

Destination Entry using Voice and Manual Methods

Investigators: Omer Tsimhoni and Paul Green

The purpose of this study is to assess the safety and usability of a couple of methods for entering complex data in moving vehicles. You will drive the simulator while entering street addresses to a computer using your voice or a touch-screen keyboard.

First, you will practice driving the simulator. There is a small risk of some motion discomfort while driving the simulator. If you feel discomfort of any significance, please let the experimenter know, so the study can be stopped. After the driving practice, you will practice entering addresses while parked and then you will enter addresses several times while driving.

For some drivers, these tasks are quite difficult. Regardless of the difficulty, it is important to try to do your best on both tasks.

Your driving will be videotaped for future reference to your data-entry performance, driving performance, and reactions overall. This study is not intended to be a test of your skill, but rather how well the system has been designed to suit you.

The study should take about two and a half hours, with a five-minute break scheduled in the middle. You will be paid $35 for your time. You can withdraw from this study at any time without penalty.

I specifically agree to be videotaped in this study and understand that (1) a copy may be provided to the sponsor and (2) selected segments from the tapes may be used in presentations to explain the results. The raw tapes will be erased 10 years after the project is completed.

Sign your name ____________________

I HAVE READ AND UNDERSTAND THE INFORMATION PRESENTED ABOVE. MY PARTICIPATION IN THIS STUDY IS ENTIRELY VOLUNTARY.

____________________ ____________________

Print your name Date

____________________ ____________________

Sign your name Witness (experimenter)

APPENDIX F. Biographical Form

Date: ________ Participant number: ________

Destination Entry Using Voice and Manual Methods – Biographical Form

Personal Details

Name ____________________________

Born (month / day / yr) ___/___/___ in (city / state) ________________

Handedness (circle one) left right

Phone: ____________________________

Email address _________________________

- May we email you for future studies? Yes No

Education

(circle highest level completed, fill in blank)

high school

some college/major :________________

college degree :________________

graduate school: major _______________

Occupation: ________________

Are you a native English speaker? (circle one) Yes No

Driving

What motor vehicle do you drive most often?

Year: _________________ Make: _______________ Model: _______________

How many miles do you drive per year? ____________

How much time do you spend on an average day driving (not as a passenger?)

__________ hours

Have you driven more than 30,000 miles in your lifetime? Yes No

Do you have any special driving licenses (e.g. heavy truck) and if so, what kind?

No Yes: explain -> __________________

How many accidents have you been involved in during the past 5 years? _____

In-Vehicle Navigation

Have you ever used an in-vehicle navigation system? No Yes

If yes: Once A few times Many times I own/owned a system

Typing Skills

How many hours a day do you type on a keyboard? __________

At what age did you learn typing? __________

Vision Circle what vision correction you use

When driving: no-correction contacts glasses (multifocal, bifocal, reading, far-vision)

When reading: no-correction contacts glasses (multifocal, bifocal, reading, far-vision)

For the experimenter only

12526616

Far Acuity	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	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
Near Acuity	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	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

	Keyboard	Touch screen
Typing speed:	__________	__________
Typing accuracy:	__________	__________

Comments:

APPENDIX G. Post-Test Evaluation Form

Date: ________   Participant number: ________

Post-test Evaluation Form

Please fill numbers (from 1 to 10) in the highlighted boxes according to the instructions
(1=strongly disagree, 10=strongly agree)
(1=extremely easy, 10=extremely difficult)

Difficulty: *(read all 3 questions first)*	Keyboard	Spelling	Words
1. It was difficult to enter addresses while parked			
2. It was difficult to enter addresses while driving			
3. It was difficult to drive while entering addresses			

(1=strongly disagree, 10=strongly agree)

Safety:	Keyboard	Spelling	Words
4. It is unsafe for me to enter addresses while driving			
5. Drivers should not be allowed to enter addresses while they drive			

(1=strongly disagree, 10=strongly agree)

General:	Keyboard	Spelling	Words
6. If this system was installed in my car I would consider using it while driving			
7. If this system was installed in my car I would consider using it while parked			

Explain: ____

Comments about this study? (please think of at least 2 ...)

____

____

APPENDIX H. Modified Cooper-Harper Scale

	Difficulty	Operator Demand Level	Rating
Mental workload is acceptable.	Very Easy Highly Desirable	Operator Mental Effort is Minimal and Desired Performance is Easily Attainable	1
	Easy, Desirable	Operator Mental Effort is Low and Desired Performance is Attainable	2
	Fair, Mild Difficulty	Acceptable Operator Mental Effort is Required to Attain Adequate System Performance	3
Mental workload is HIGH and should be reduced. Errors are small and inconsequential.	Minor But Annoying Difficulty	Moderately High Operator Mental Effort is Required to Attain Adequate System Performance	4
	Moderately Objectionable Difficulty	High Operator Mental Effort is Required to Attain Adequate System Performance	5
	Very Objectionable But Tolerable Difficulty	Maximum Operator Mental Effort is Required to Attain Adequate System Performance	6
Instructed task can be accomplished most of the time. Errors are large or consequential.	Major Difficulty	Maximum Operator Mental Effort is Required to Bring Errors to Moderate Level	7
	Major Difficulty	Maximum Operator Mental Effort is Required to Avoid Large or Numerous Errors	8
	Major Difficulty	Intense Operator Mental Effort is Required to Accomplish Task, But Frequent or Numerous Errors Persist	9
Task is impossible to accomplish.	Impossible	Instructed Task Cannot Be Accomplished Reliably	10

APPENDIX I. Sample Text for Typing Test

This sample text was given to test typing speed in a 1-minute interval. The text consists of 100 words with 4.1 mean characters per word and a Flesch-Kincaid grade level of 7.7.

The famous sea explorer James Cook was born on 28 October 1728 in a village in Yorkshire in England. He was the son of a poor Scottish farmer and thus was very unlikely to have seen many books before going to school. There he is said to have been very good at arithmetic.

At the age of twelve James left his home in order to learn some shopkeeper's trade in a fishing village. That must have been quite a change in the young boy's life as he had probably never left his home village before. On his way to the

APPENDIX J. Observed Noise in Lateral Lane Position

The lateral lane position, sampled at 30 Hz, appeared with noise that was characterized as a continuous wave at 2 - 2.5 Hz and an amplitude of 0.06±0.03 m (0.2±0.1 ft). (See Figure 54.) Further investigation of this phenomenon revealed that the noise appeared only on curves. Lateral position and yaw angle were affected, but steering wheel angle was not.

The explanation to this effect lies in the design of curves in the simulator. Every road in the simulator is a series of straight segments of 30 ft (9.1 m), which may be connected at an angle to form a curve. When a curve is driven, the steering wheel angle is typically held at a constant position and the curve is driven as it would on a perfect curve. However, the lateral position in the lane is measured relative to the actual center of the lane, which is segmented. This leads to a measurement error which is potentially greatest at the center of every segment (after 15 ft from the beginning of the segment) and smallest at the beginning and end of each segment.

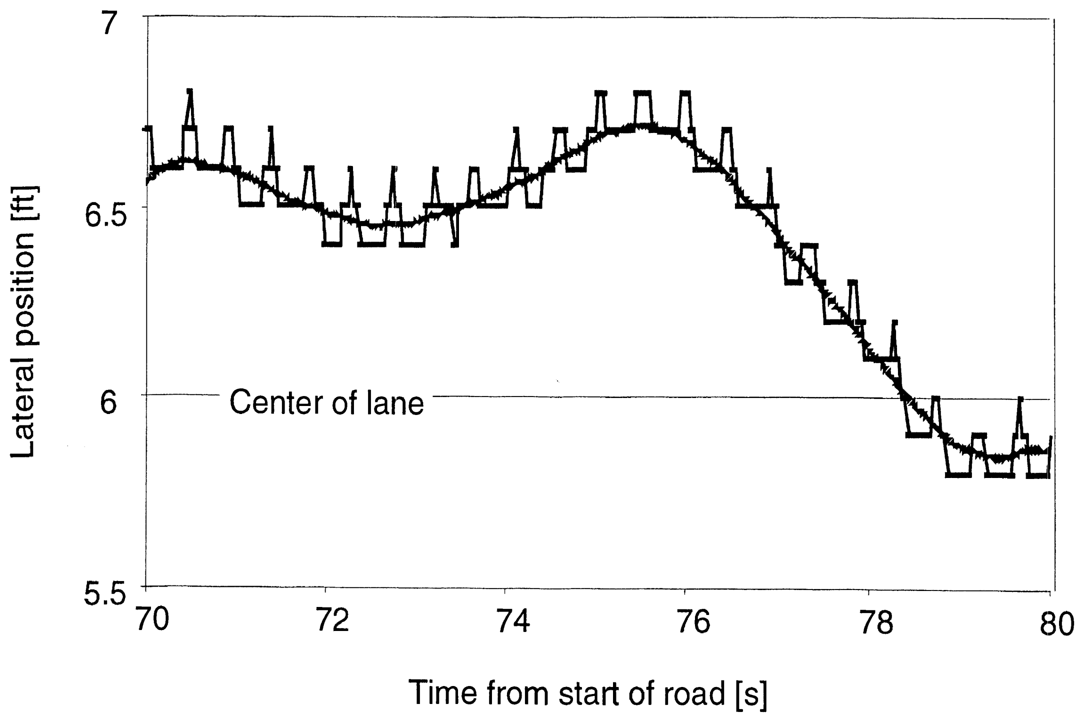


Figure 54. Recorded lateral position for a sample of a curve

Figure 55 illustrates the maximum lateral distance (E) between an actual road segment and a perfect curve. At a curve radius of R=194.1 m and segment length of Y=9.1 m, E = 0.054 m. The magnitude of this error matches the observed amplitude of the noise.

The frequency of the noise can be explained by the forward velocity of the car. At 45 mi/h, the car passes 2.2 segments per second, which lies within the observed range of the noise frequency (2 - 2.5 Hz)

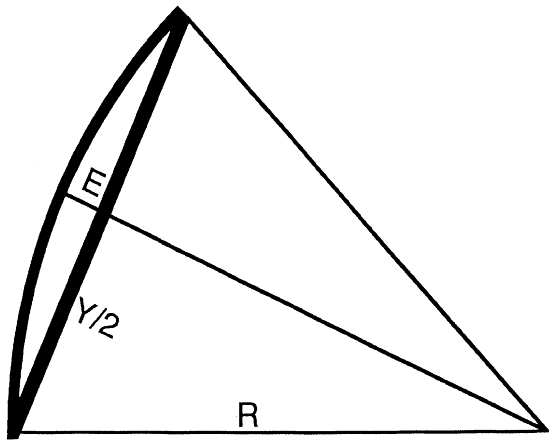


Figure 55. A lateral error (E) is caused by the segmentation of the curve

To eliminate this artifact, the lateral lane position in the graphs was smoothed using a running average with a window of 1.5 s.