Driving Performance Evaluation of the Ali-Scout Navigation System: A Preliminary Analysis

Paul Green
pagreen@umich.edu
University of Michigan Transportation Research Institute
Human Factors
2901 Baxter Rd. Ann Arbor, Michigan 48109-2150 USA
Phone: (734) 763-3795 Fax: (734) 764-1221

Jill Fleming
jsf@umich.edu
University of Michigan Transportation Research Institute
Human Factors
2901 Baxter Rd. Ann Arbor, Michigan 48109-2150 USA
Phone: (734) 763-2498 Fax: (734) 764-1221

Stewart Katz
stewkat@umich.edu
University of Michigan Transportation Research Institute
Human Factors
2901 Baxter Rd. Ann Arbor, Michigan 48109-2150 USA
Phone: (734) 763-2498 Fax: (734) 764-1221

ABSTRACT

This paper describes the first of two experiments examining the safety and usability of the Siemens Ali-Scout navigation system, an in-vehicle interface that provides turn-by-turn visual and voice guidance. A total of 54 drivers varying in age drove an instrumented car to four destinations, twice using the Ali-Scout and, subsequently, once using experimenter verbal guidance. Subjects were tested in the afternoon, at rush hour, and in the evening. There were no crashes or near misses using the Ali-Scout, but there were four critical incidents where drivers changed lanes in response to navigation voice instructions without checking traffic. Excluding the turns into destinations, the turn error rate was 8 percent, and uncertainties occurred at an additional 13 percent of the turns. Most of the errors and uncertainties occurred in autonomous mode in which only the distance and direction to the destination were shown, not turn-by-turn guidance. Driving performance measures summarized in this paper include the mean speed while moving, and the standard deviations of throttle, speed, and lateral position. The first three of these measures were more sensitive to interface differences than the standard deviation of lateral position, a measure of lane variability. Generally, there were significant differences due to age, subject, destination, and section within destination for these measures. Subject differences were quite pronounced, with variability measures for the best and worst subject in each age-sex group differing by 2:1 to 3:1. The results suggest that the collection of variables sensitive to navigation interfaces is large, posing challenges for the development of a simple standard safety assessment.

INTRODUCTION

Navigation systems can make driving safer by reducing the need to look at paper maps or written directions while driving, allowing drivers to maintain their attention on the road. Navigation systems make driving more convenient (for prospective customers) and allow people to travel with confidence. From the government's perspective, navigation systems support more efficient use of the existing road network, as well as reducing air pollution and wasted fuel, all desired outcomes.

There are two primary driver-related navigation tasks: (1) entering and retrieving destinations, and (2) following the directions given by these systems (route guidance). Destination entry has been covered elsewhere (Steinfeld, Manes, Green and Hunter, 1996) and will be covered in a forthcoming report (Manes, Green, and Hunter, 1997). This paper covers driver performance while using route guidance, providing selected results from a lengthy report on that topic (Katz, Fleming, Green, Hunter, and Damouth, 1996).

Several previous studies have examined similar issues concerning driver performance with navigation systems (Parkes, Ashby, and Fairclough, 1991; Daimon, 1992; Green, 1992; Green, Williams, Hoekstra, George, and Wen, 1993; Green, Hoekstra, and Williams, 1993; Dingus, McGehee, Hulse, Jahns, Manakkal, Mollenhauer, and Fleischman, 1995; Daimon and Kawashima, 1996; Foley and Hudak, 1996). Key findings from previous studies include the following.

1. Turn errors have indicated interface differences in every study in which they have been used.

2. Input measures of longitudinal and lateral control (steering wheel variance, throttle variance, etc.) have not proven to be sensitive to interface differences, possibly because of measurement errors.

3. Of the longitudinal output measures, both mean speed and speed variance (along with trip duration) are sometimes sensitive to some interface differences.

4. Of the lateral output measures, the standard deviation of lane position seems to be more sensitive than the number of lane excursions, mainly because there are often so few excursions in real experiments.

Issues Examined

This project, part of an FHWA-sponsored operational field test, was intended to give generalizable results concerning driver behavior and performance while using the Ali-Scout navigation system. Keeping this goal in mind and the gaps in the literature, the following issues emerged.

1. What were typical values for driving performance and behavior while using the Ali-Scout interface? How did those values change with driver age, sex, time of day, traffic, and experience with the system?

2. Were there safety and usability problems associated with the Ali-Scout interface?

This paper emphasizes the driving performance results from the first of two experiments. Subjective responses and additional analyses of the performance data appear in Katz, Fleming, Green, Hunter, and Damouth (1996).

TEST PLAN

Ali-Scout Operation

The Siemens Ali-Scout navigation system provides directions to destinations using auditory cues and a visual display installed in the user's vehicle. A computer system at the traffic operations center communicates the best route for each vehicle via line-of-sight infrared beacons. To reduce cost, in-vehicle map databases and GPS receivers were not utilized. Accordingly, vehicles equipped with the Ali-Scout must periodically encounter beacons for navigation updates, a limiting feature. Beacons were located at major intersections (approximately 1 mile apart) and on highways in Oakland County, Michigan, a suburb of Detroit. Of the 18 beacons operational when this experiment was conducted, 5 were located along the test route.

The Ali-Scout has two guidance modes: (1) autonomous, in which a "crow-fly" distance (in miles) and directional arrow are provided (indicated by an A in the upper left corner of the display), (2) guided, in which turn arrows and turn-by-turn voice guidance are provided. (See Figure 1.)

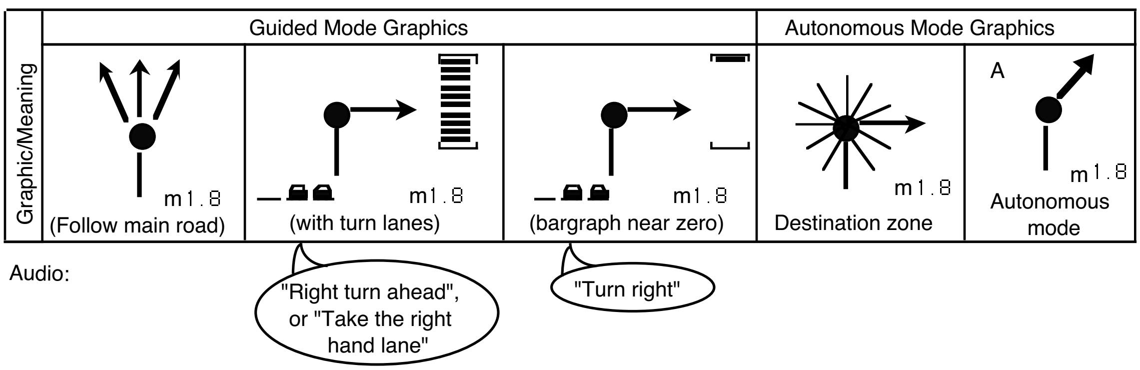

Figure 1. Ali-Scout maneuver graphics and auditory messages.

Generally, the Ali-Scout begins in autonomous mode, providing only a crow-fly arrow as guidance. Once a beacon is passed, the system will guide drivers with auditory and visual cues to within a quarter mile of the desired destination, at which time the unit switches back into the autonomous mode.

Test Activities and Their Sequence

All subjects used the Ali-Scout interface. Subjects were tested at one of three times: afternoon (2 PM) associated with light to moderate traffic, rush hour (5 PM) associated with heavy traffic, and in the evening (9 PM) associated with little to no traffic. The experiment consisted of two sessions, one week apart. During trial one, the subject drove a fixed route, consisting of four destinations, with the aid of the navigation system only. (See Table 1.) Upon returning for session two, the subject completed two additional trials of the same route. The first trial of session two was identical to that of trial one; the subject used only the navigation system to arrive at each destination. During the second trial (baseline) of session two, subjects were given turn by turn directions by the experimenter. Subjects were free to ask questions concerning where and when they should turn at any time. This baseline provided comparison data on how subjects would drive to a reasonably familiar destination when provided close to ideal guidance (voice instructions from the experimenter). Driving differences between Ali-Scout guided trials and the experimenter-guided trial should reflect the workload associated with using the Ali-Scout unit. Differences between Ali-Scout guided trials should provide insight about the learning aspects of the interface.

Table 1. Test route description.

Destination Number/ Name	Turns/Maneuvers	Road Description (# of lanes)	Speed Limit (mi/hr)	Traffic
1. SOC Credit Union	•verbal instructions to I-75 N. •guided right onto exit ramp •guided left onto Crooks Rd. •autonomous right into destination	•I-75: 3 •Exit ramp: 1 •Crooks: 3 merges to 2	•65 •25 •45	•heavy •light •heavy
2. Harlan Plaza	•verbal instructions to Crooks Rd. •autonomous right onto Long Lake •guided right onto Rochester Rd. •guided left onto Wattles Rd. •autonomous left into destination	•Long Lake: 2 merges to 1 •Rochester: 2 •Wattles: 1	•45 •45 •40	•heavy •heavy •moderate
3. Cumber-land Dr.	•autonomous left out of parking lot •autonomous right onto John R Rd. •autonomous right onto Cumberland Dr.	•Wattles: 1 •John R: 2 •Cumberland: 1	•40 •45 •25	•moderate •moderate •residential
4. Maplewood Plaza	•verbal instructions to Wattles Rd. •guided left onto Rochester Rd. •guided right onto Rochester Rd. split •autonomous right into destination	•Wattles: 1 •Rochester: 2 •Rochester Rd. split: 2	•40 •45 •35	•moderate •heavy •moderate

Test Participants

A total of 54 licensed drivers (18 young, 18 middle-aged, 18 older) were tested. Each age group (means of 21, 48, and 72 years, respectively) contained an equal number of men and women. Corrected visual acuity ranged from 20/15 to 20/30. Only one subject had previously driven a vehicle with a navigation system, an UMTRI prototype (Green, Williams, Hoekstra, George, and Wen, 1993).

Subjects were either friends of the experimenters, respondents to a newspaper advertisement, or patrons of a senior center. All samples reflected subjects who were moderately familiar with the city of Troy (approximately 3.0 on a scale of 1 (not at all familiar) to 5 (very familiar)). Driving experience ranged between 12,000 and 16,000 miles/year.

Instrumented Car

Data was collected using an instrumented, left-hand drive 1991 Honda Accord station wagon. This car has sensors for all major driver inputs (steering wheel angle, throttle and brake position, turn signals), vehicle responses (speed, lateral position), the headway distance to a lead vehicle, and cameras for recording the forward scene and driver.

The video recording system consists of three bullet (lipstick) cameras (one to record the forward scene mounted below the inside rear view mirror, a second aimed at the driver and mounted on the left side A-pillar, a third aimed at the driver and mounted below the Ali-Scout navigation unit to record eye fixations), and two small cameras located in the outside mirrors to record the lane markings on either side of the vehicle (lane trackers). Camera outputs are combined, along with a summary of the data collected by a computer, by a quad splitter, displayed on a monitor, and recorded on a VCR. The two lane tracker images are combined by a two-image splitter and fill one quadrant of the quad splitter image. (See Figure 2.) Sound was also recorded using miniature microphones.

Figure 2. Typical quad-screen image.

RESULTS AND DISCUSSION

This paper only presents a sample of the results, emphasizing crash-related data, turn errors, one measure of overall driving speed (mean moving speed), two measures of longitudinal control, one input, one output (standard deviation of throttle position, standard deviation of speed (while moving)), and one measure of lateral control (standard deviation of lane position). Subject comments and questionnaire results, as well as the analysis of other performance measures appear in the Katz, Fleming, Green, Hunter, and Damouth (1996).

Crashes, Critical Incidents

Fortunately, no crashes were experienced during experimentation. However, several critical incidents did occur. Critical incidents were defined as situations when something drastically wrong occured (passing a stop sign without slowing or noting it was there) but because of the traffic situation (no vehicles were present), no harm occured. During the initial exposure to the Ali-Scout (trial one), four critical incidents occurred when drivers were given an auditory command to either take the right- or left-hand lane. Four drivers responded to the command by immediately changing lanes, neglecting to check their blind spot and ignoring surrounding traffic. In response, other drivers honked at the test vehicle, causing the subject to swerve back into the original lane, without further incident. No critical incidents were experienced during the second trial with the Ali-Scout, indicating that with practice these messages may not be perceived as commands.

Turn Errors and Uncertainties

A destination was missed when the subject drove by the destination without noticing it, or when identifying the destination too late to safely make the turn. Table 2 shows missed destinations by trial. There were no errors during the experimenter-guided trial (baseline). Notice, that the largest number of errors occurred for the first trial and the first destination, indicating that learning was a factor.

Table 2. Destinations missed by subjects while using Ali-Scout.

Destination	Number of subjects to miss destination	
	Trial 1	Trial 2
1. SOC Credit Union	27	11
2. Harlan Plaza	4	1
3. Cumberland Dr.	3	3
4. Maplewood Plaza	14	2

Turn errors occurred when subjects turned too soon or passed intersections where they should have turned. Turn uncertainties involved either erratic turns or expressions of uncertainty about what to do. ("I do not know if I am supposed to turn here.") Table 3 shows the turn error and uncertainty data. The turn error rates were 7.5 percent for the first trial and just over 2 percent for the second. However, when uncertainties are included, the combined rates are 10.5 percent and 5 percent for the first and second trials respectively. Examination of data show that about 3/4 of the errors and uncertainties occurred in autonomous mode even though only a small fraction of the driving occurred in that mode. This suggests a significant disadvantage for interfaces that only give a distance and a bearing (such as the Delco Telepath).

Table 3. Ali-Scout turn errors and uncertainties.

Route to Dest.	Ali-Scout Mode	Error Description	Number of turn errors by trial		Number of uncertainties by trial	
			1	2	1	2
2	A	Missed right turn at Long Lake	20	7	16	12
2	G	Turned into street prior to correct turn	2	0	3	0
2	G	Missed left turn at Wattles	3	1	8	4
2	G	Turned before intersections into shopping plaza	1	0	1	1
3	A	Turned right instead of left out of parking lot	3	1	5	0
3	A	Missed left turn onto John R	1	1	10	3
4	G	Missed right turn onto Rochester	2	0	3	2
		Total	32	10	46	22

Note: A=Autonomous, G=Guided

By way of comparison, Green, Williams, Hoekstra, George, and Wen (1993) reported error rates of 3, 2, and 1 percent for auditory, instrument panel, and HUD-based navigation systems respectively, each driven by 10 drivers over a more difficult 19-turn test route. These error rates were far less than those for the Ali-Scout. This suggests the Ali-Scout was much less useful than the UMTRI prototype interface (and, in fact, as found in a follow-up study, also more error prone than the Rockwell PathMaster). Further, Green, Williams, Hoekstra, George, and Wen (1993) found that the number of near miss errors, uncertainties, and/or confusions at intersections was about 50 percent greater than the actual number of turn errors, comparable to the values reported here.

For the Ali-Scout evaluation, navigation problems varied little with the age and sex of the driver. There were a total of 24, 19, and 25 instances of turn uncertainties for young, middle-aged, and older drivers respectively. Hence, from the perspective of navigation efficiency, this experiment suggests age and sex need not be considered in future work.

Driving Data Reduction and Analysis

The driving data were reduced via custom software that added codes to the data set and partitioned it into sections. Of the 864 route sections examined over all subjects, 41 were lost due to equipment, software, or experimenter error, and those cases were randomly distributed through the data set. Each dependent measure of interest was examined using ANOVA. Terms in the model included subject-related effects and their interactions (age group, sex, subject nested within sex and age, and the sex by age interaction), trip characteristics (destination, road section within destination, session time), navigation system differences (trial) and selected interactions likely to be significant (destination by session time, destination by trial number). Only selected measures will be reported here.

Mean Speed While Moving

Because of speed sensor limitations, moving speed was defined as any time the speed was above 3.5 mi/hr. The advantage of looking at mean moving speed, as opposed to overall mean speed, is that it eliminates the randomness associated with waiting at traffic lights. In the ANOVA of this measure, there were significant differences due to age, subjects, destinations, and trial number (all p=0.0001). Within subjects, the primary difference was that older drivers tended to drive more slowly (36.2, 36.0 and 34.6 mi/hr for young, middle, and older subjects). Figure 3 shows the size of the individual differences (though confounded with time of day effects since subjects drove at three different times).

The session time was found to be significant (p=0.0057). The mean moving speed for the afternoon session was 36.1 mi/hr, rush hour session was 35.0 mi/hr, and night session was 35.7 mi/hr. The mean was lowest during the rush hour session probably due to the high traffic density.

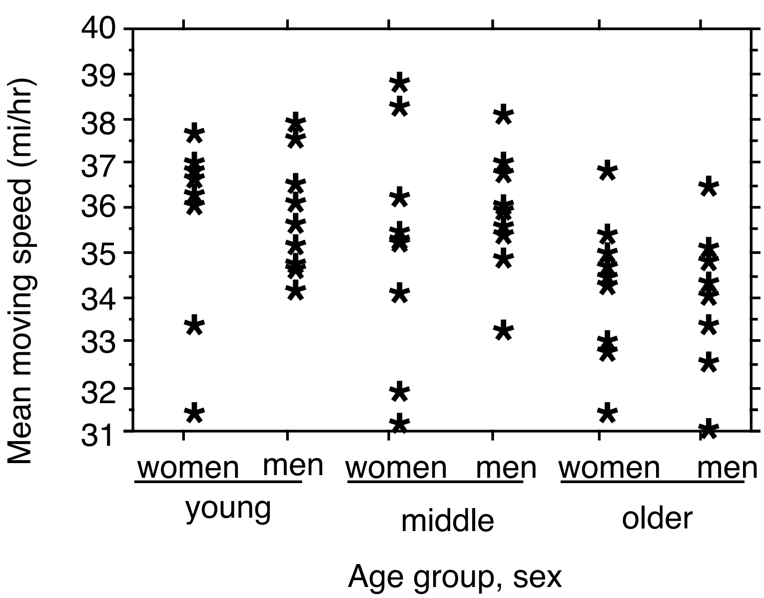

Figure 3. Mean moving speeds for each subject.

The driving trial effect was significant (p=0.0001). Trials one, two and three (baseline) had mean moving speeds of 34.8 mi/hr, 35.4 mi/hr, 36.6 mi/hr respectively. The difference between trial one and two was 0.6 mi/hr, whereas the difference between baseline and trial two was 1.2 mi/hr. Thus, while the effects of practice with the Ali-Scout only led to a marginal increase in moving speed, ideal guidance increased mean moving speed significantly more.

Speed Standard Deviation While Moving

Age had a significant effect on the speed standard deviation (p=0.0001) with mean standard deviations of 8.7, 8.6, and 8.2 mi/hr for young, middle-aged, and older drivers respectively. Individual subject means ranged between 6.8 and 9.8 mi/hr.

Speed variance was influenced by two factors, overall speed (variance increases with the mean), and driver erraticness (more erratic driving leads to greater speed variability). In this case, it appears the effect of mean speed predominated (leading to greater variance). Destinations 2, and 4 have greater mean speed limits than destination 3 (see Figure 4). This does not hold for destination one, however, as it contained a section of highway where there were no traffic lights, not the case for the other three destinations. The differences found, while significant, are so small that they are of minimal practical significance. This suggests that due to confounding, speed variance may not be an optimal measure of driving safety in navigation studies.

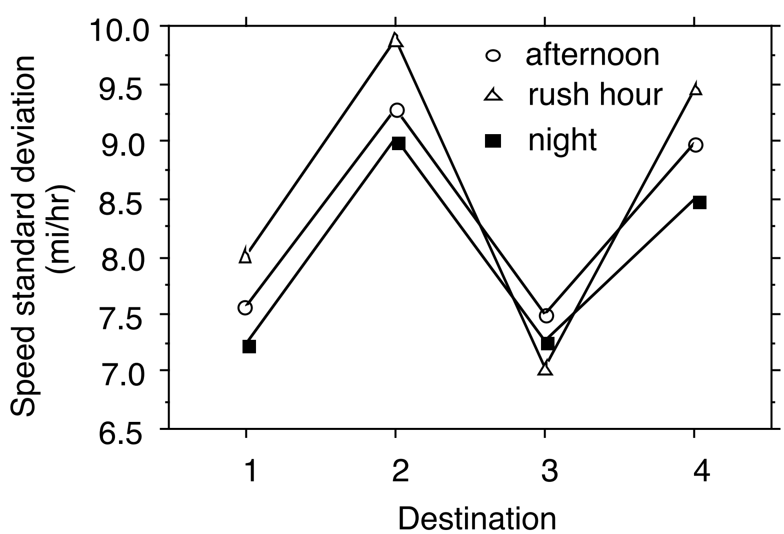

Figure 4. Mean standard deviation of speed by destination.

The time at which test sessions were run was found to be significant (p=0.0001). The mean speed standard deviations for the afternoon, rush hour, and night sessions were 8.5, 8.9, and 8.2 mi/hr respectively. The larger traffic density present at rush hour, and during the afternoon was a contributing factor to their increased mean standard deviation of speed.

The trial number was also significant ($p=0.0002$). The mean speed standard deviations for trial one, trial two, and the baseline trial were 8.3, 8.7, and 8.6 mi/hr respectively. Thus, practice with the Ali-Scout interface and use of experimenter guidance increased speed variability.

Standard Deviation of Throttle Position

The variability of throttle position, how much the driver changes the position of the gas pedal from moment to moment, is measured in percent. The standard deviation of throttle position is the input causing speed to vary.

Interestingly, there were no age or sex differences, whereas there were age differences in speed variability. However, there were significant differences ($p<0.0001$) due to destination, section within destination, session time, trial number, and subjects. Individual subject means ranged from 4.5 to 11.5 percent. The mean standard deviation of throttle position for trials one, two, and baseline were 6.8 percent, 7.3 percent, and 7.6 percent respectively. A counter-intuitive result was that there was more variability in throttle position as the driver became more familiar with the route, possibly reflecting more aggressive driving. This is consistent with the speed variance results.

Standard Deviation of Lateral Position

Standard deviation of lateral position (measured in feet) is a measure that reflects how well the driver is controlling the vehicle. Greater lane variability reflects poorer driving and may lead to crashes. There were no significant differences in lane variability due to age, but sex was found to be significant ($p=0.0004$). The mean standard deviation was 0.87 ft for women and 0.91 ft for men. Individual subject differences ranged from just over 0.6 to 1.15 ft. There were also significant differences due to subject, destination, section within destination, and session time (all $p<0.0001$), but not trial number.

Figure 5 shows session time differences. Notice that at night lane variance was actually less than in the day time when it was easier to see. It could be that during the daytime session traffic (heavier at that time) was a more important factor than visibility.

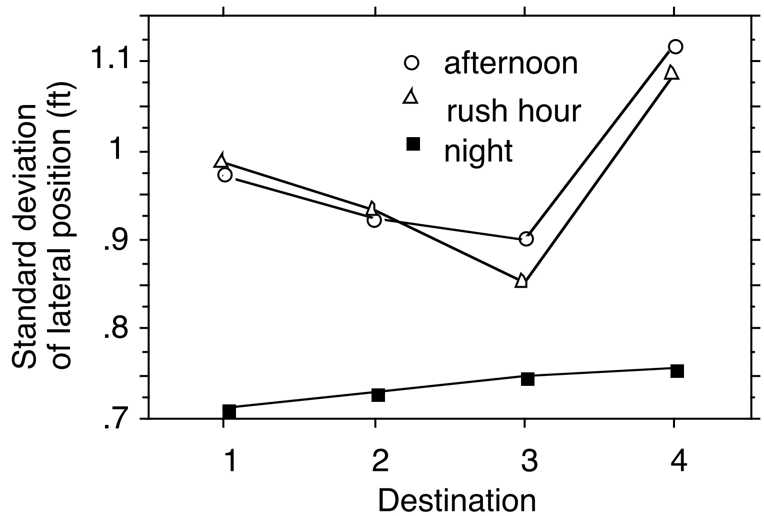

Figure 5. Mean standard deviation of lateral position for destination by session time.

The lack of a trial number effect was quite interesting as one would have thought that practice and using ideal guidance would reduce lane variance since the driver would be allowed to pay more attention to driving. (The means for trials one, two and baseline were 0.89 ft, 0.91 ft, and 0.88 ft, respectively.) These data suggest that lane-keeping (a safety measure) did not depend on the quality of guidance provided (within the range explored).

SUMMARY

While there were no crashes in this experiment, the four critical incidents during early use of the Ali-Scout navigation system raise concerns about its safety. The major problem was that subjects interpreted maneuver guidance as commands. This problem may be remedied by rephrasing voice instructions. This concern is not unique to the Ali-Scout interface and is a matter to which interface designers should pay special attention.

The turn error data showed significant differences due to practice in using the Ali-Scout interface. Performance with the Ali-Scout was not nearly as good as with the UMTRI interface. Ali-Scout users had consistent navigation problems whenever they were in the autonomous ("crow fly") mode, suggesting that such guidance systems are much less useful.

In terms of the driving data, road-related differences (destination, session within destination) were consistently significant, and in most cases, so were differences due to traffic/time of day. Of the subject-related factors, there were consistent differences due to age and some differences due to sex, however the largest differences were found among individual subjects. Within each 18 subject group, differences in variability between the best and worst subject were typically 2:1 to 3:1. Trial differences, reflecting differences in interface design and practice, were significant for mean speed (while moving), standard deviation of speed (while moving), and standard deviation of throttle position. However, the standard deviation of lateral position, a traditional measure of safety, was not significant. Differences were typically on the order of a few percent. Providing better guidance caused drivers to proceed more quickly and added to their speed variability rather than reducing it.

These results suggest that the interpretation of driving measures is much more complex than it appears at first glance. For example, navigation systems should make traffic flow both more quickly and smoothly. As a consequence, one would expect many means to increase and measures of variability to decrease. However, because the two sets of measures are correlated (for example, mean speed and speed variance), they both increase, not a desired result.

These results also indicated the need to focus on both navigation errors and navigation uncertainties. There were 50 percent more uncertainties than errors. Their combined total may be sufficient as to provide insights into navigation system differences in even more modest studies. However, because such studies are done by different investigators over different routes, a set of standard scoring rules and methods to adjust for route difficulty are desired to facilitate between-study comparisons.

It is unfortunate that the era of federally-sponsored operational field tests has come to an end. There still is significant uncertainty as to what to measure in an on-road test of a navigation system. This situation will create significant problems for those now faced with writing standardized safety test protocols both in government and industry.

REFERENCES

Daimon, T. (1992). Driver's Characteristics and Performances When Using In-vehicle Navigation System, Proceedings of the Third International Conference on Vehicle Navigation and Information Systems, New York: Institute of Electrical and Electronics Engineers, 251-261.

Daimon, T. and Kawashima, H. (1996). New Viewpoints for Evaluation of In-vehicle Information Systems: Applying Methods in Cognitive Engineering, JSAE Review, 17, 151-157.

Dingus, T., McGehee, D., Hulse, M., Jahns, S., Manakkal, N., Mollenhauer, M., and Fleischman, R. (1995). Travtek Evaluation Task C3 - Camera Car Study (Technical Report FHWA-RD-94-076), McLean, VA: U. S. Department of Transportation, Federal Highway Administration.

Foley, J.P. and Hudak, M.J. (1996). Autonomous Route Guidance System Field Test, Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting, Santa Monica, CA: Human Factors and Ergonomics Society, 887-890.

Green, P. (1992). American Human Factors Research On In-Vehicle Navigation Systems (Technical Report UMTRI-92-47), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P., Hoekstra, E., and Williams, M. (1993) On-The-Road Tests of Driver Interfaces: Examination of a Navigation System and a Car Phone (Technical Report UMTRI-93-35), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P., Williams, M., Hoekstra, E., George, K. and Wen, C. (1993). Initial On-the-Road Tests of Driver Information System Interfaces: Examination of Navigation, Traffic Information, IVSAWS, and Vehicle Monitoring (Technical report UMTRI-93-32), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

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

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

Parkes, A.M., Ashby, M.C., and Fairclough, S.H. (1991). The Effect of Different In-Vehicle Route Information Displays on Driver Behaviour (SAE paper 912734), Vehicle Navigation and Information Systems-1991 (VNIS'91), Warrendale, PA: Society of Automotive Engineers, 61-70.

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

List of Tables and Figures by title