Address Entry While Driving: Speech Recognition Versus a Touch-Screen Keyboard

Omer Tsimhoni, Daniel Smith, and **Paul Green,** University of Michigan Transportation Research Institute, Ann Arbor, Michigan

A driving simulator experiment was conducted to determine the effects of entering addresses into a navigation system during driving. Participants drove on roads of varying visual demand while entering addresses. Three address entry methods were explored: word-based speech recognition, character-based speech recognition, and typing on a touch-screen keyboard. For each method, vehicle control and task measures, glance timing, and subjective ratings were examined. During driving, word-based speech recognition yielded the shortest total task time (15.3 s), followed by character-based speech recognition (41.0 s) and touch-screen keyboard (86.0 s). The standard deviation of lateral position when performing keyboard entry (0.21 m) was 60% higher than that for all other address entry methods (0.13 m). Degradation of vehicle control associated with address entry using a touch screen suggests that the use of speech recognition is favorable. Speech recognition systems with visual feedback, however, even with excellent accuracy, are not without performance consequences. Applications of this research include the design of in-vehicle navigation systems as well as other systems requiring significant driver input, such as E-mail, the Internet, and text messaging.

INTRODUCTION

Motor vehicle manufacturers expect that in the near future, a significant share of their profits will be associated with the sales of telematic devices – that is, computer-based in-vehicle information and communication systems such as cell phones and navigation systems (Richardson & Green, 2000). There is concern, however, that using such devices may overload drivers and increase crash risk. This includes visual-manual interfaces that capture driver attention and induce drivers to look at the road less often (e.g., Wierwille & Tijerina, 1998) and auditory-speech interfaces, such as phone systems, for which the cognitive demands of conversation reduce awareness of the driving situation (e.g., Strayer, Drews, & Johnston, 2003).

Of the telematic tasks of concern during driving, entering a street address into a navigation system is often mentioned, and manual entry has been the topic of several studies (e.g., Chiang, Brooks, & Weir, 2001; Farber et al., 2000; Nowakowski, Utsui, & Green, 2000; see Green & Shah, 2003, for a review.)

Speech input is believed to be the ideal modality for information entry because it should present less competition for visual resources. However, speech interfaces with auditory feedback are not without cost. Lee, Caven, Haake, and Brown (2001), for example, found a significant delay in response time to braking events of a lead vehicle when drivers interacted with an auditory-speech system. In one of the few studies comparing interface modalities for destination entry, Tijerina, Parmer, and Goodman (1998) found that a particular character recognition voice interface had shorter task times, required fewer glances, and had shorter mean glance durations than did several other manual interfaces.

Speech interfaces based on word recognition and the effect of recognition accuracy on driver performance have not been examined. At first thought, word-based recognition would seem

Address correspondence to Omer Tsimhoni, Human Factors Division, University of Michigan Transportation Research Institute, 2901 Baxter Rd., Ann Arbor, MI 48109-2150; omert@umich.edu. **HUMAN FACTORS,** Vol. 46, No. 4, Winter 2004, pp. 600–610.

ideal. In real systems, however, errors will occur and feedback needs to be provided to the driver. Often, for reasons of cost, speed, and technology, feedback is provided visually, leading to unknown visual and cognitive demands that compete with the primary task of driving.

To make cost-benefit engineering decisions regarding interface design, these aspects of speech recognition systems should be quantified. The current experiment considered them by comparing speech recognition of addresses with entry of the same addresses using a touch-screen keyboard. Speech recognition was split into word-based and character-based recognition. The latter method imitates keyboard entry using speech instead of key presses, allowing for a better comparison of the difference between speech input and motor input. Recognition accuracy was experimentally controlled and fixed across participants to provide some insights regarding the effects of errors on performance. Finally, driving performance was evaluated at the vehicle control level. For that kind of analysis, using curves of varying curvatures has been shown to be an effective method to keep the visual demands of the driving task steady (Tsimhoni & Green, 2001).

The objective of this experiment was to compare the effects of manual and voice entry methods on task performance and vehicle control as a function of several controlled levels of visual demand of driving. (Note: See Tsimhoni, Smith, & Green, 2001, the report on which this paper was based.)

METHOD

Participants

Twenty-four licensed drivers participated in this experiment, 12 younger (age 20–29 years, mean = 24) and 12 older (age 65–72 years, mean = 69), with equal numbers of men and women in each age group. Participants were recruited via an advertisement in the local newspaper and were paid $40. All participants had far visual acuity of 20/40 or better. All had midrange (80 cm) visual acuity of 20/70 or better and no color deficiencies. All older participants were retired but had maintained active lifestyles. Prescreening of all participants ensured they had good driving records and were physically healthy.

Experimental Design

Each participant used three entry methods (word-based speech recognition, character-based speech recognition, and typing on a touch-screen keyboard) combined with four levels of driving workload (parked, straight, moderate curves, and sharp curves). Because of a lack of effects observed in pilot tests and to keep the duration of the experiment reasonable, 3 of the 12 combinations were not tested: word-based speech recognition on straight sections and on moderate curves, and character-based speech recognition on moderate curves.

Participants in each age-gender subgroup were randomly assigned to one of two groups, which performed either the keyboard task followed by the speech recognition task or vice versa. Within each of these groups of 3 participants, the order of curvature was manipulated following a Latin square design so that each of the three road curvatures appeared first, second, or third for exactly 1 participant.

Test Materials and Equipment

Driving simulator. The experiment was conducted in the second-generation University of Michigan Transportation Research Institute (UMTRI) driver interface research simulator, a fixed-based driving simulator based on a network of Macintosh computers (UMTRI, 2001). The simulator consisted of a cab, a projection screen, a torque motor connected to the steering wheel, a sound system, a subbass sound system for vertical vibration, and a computer system to project images of an instrument panel. The forward-view projection screen, offering a field of view of 33° × 23°, was 6 m (~20 feet) in front of the driver, effectively at optical infinity.

Simulated roads. The simulated roads had three levels of road curvature (straight sections, moderate curves of 582 m radius, and sharp curves of 194 m radius; 0°, 3°, and 9° of curvature, respectively). Both lanes of the two-lane road were 3.66 m (12 feet) wide. Traffic in both lanes was headed in the same direction and consisted of a platoon of four vehicles: a lead vehicle driving in the right lane and three additional vehicles driving in the left lane. The participant was instructed to drive in the right lane at a comfortable distance behind a lead vehicle, which maintained a constant speed of 72 kph (45 mph).

Touch-screen keyboard. A touch-screen display (Elotouch 1225L) with a 7-inch (17.8-cm) diagonal opening in a cardboard cover was used to simulate an entry module of a navigation system. The touch screen was located in the center console of the vehicle, 23° ± 2° below the horizontal line of sight and 30° ± 2° to the right of center (Figure 1).

A standard QWERTY key arrangement was selected based on Coleman, Loring, and Wiklund (1991). Keys were 12.7 mm high and 12 mm wide and were spaced 0.7 mm apart.

Memo display. Addresses to be entered appeared on a 5-inch (12.7-cm) LCD (Advanced Video and Communication Inc., Model AM-064P) located to the right of the driver at arm's length. The display simulated a low-cost address book or a piece of paper on which an address may appear. To allow easy reading, characters on the display were relatively large (character height = 11 mm, 1° at 63 cm). The memo display was 29° ± 2° below the horizontal line of sight and 56° ± 2° to the right of the center (Figure 1).

Speech recognition. The Wizard of Oz method (e.g., Green & 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 characters the participant said. This method was chosen because a high-accuracy hardware-based speech recognition system was not readily available and would have been costly. Further, the Wizard of Oz method allowed full control of the accuracy of recognition and did not require a training session, as would a commercial system, thus shortening the experiment duration.

When the speech recognition system was word based, the participant was instructed to say the address word by word, briefly pausing between address elements (city, street name, street suffix, and number). When the speech recognition system was character-based, the participant was instructed to spell out each of the four address elements. Participants were instructed to respond to system recognition errors as soon as they detected them (either after a single word or at the end of the address) by saying "scratch that" for each word they wanted to erase.

Figure 1. Driver's view of the road and touch screen.

Each experimental block included 24 items (6 addresses of 4 items each), of which 2 items (the same for all participants) were deliberately not recognized. Thus, word recognition accuracy of the system was about 92% per word, performance that is typical of contemporary systems.

To support the participants' impression that the speech recognition system was real, the preset recognition errors that the system made consisted of acoustically confusable letters (e.g., *T, P,* and *B:* "Teapody" instead of "Peabody"). Words appeared on the display 2 s after the last character for the word was uttered, mimicking the delay of contemporary speech recognizers. Consistent with this approach, a tone indicating the acceptance of the full address, or the need to redo it, was played 3 s after the last digit was uttered. Observations and participants' comments suggest that most participants did not know the interface was simulated and that they interacted with the system as if it were real.

Address list. The addresses used in this experiment were carefully chosen to be realistic, consistent, and not easily predictable. All addresses were 20 characters long, and the number of characters in city names and street names varied between 5 and 9 to provide variation in the typing process. To reduce practice effects, street names and city names did not repeat throughout the experiment. Address elements were entered in a fixed order: city, street, street suffix, and building number (e.g., "Dexter, Broadway Dr, 8709" and "Chicago, Gannett Rd, 1672").

Test Activities and Sequence

Table 1 summarizes the order of activities for a typical participant. Participants were given considerable practice before testing began and performed single tasks before multiple tasks.

RESULTS

Total Task Time

Of the three entry methods, word-based speech recognition was fastest, 14.7 s, $F(2, 40) = 70.6$, $p < .0001$ (Figure 2), and times did not increase significantly when performed while driving (14.2 s while parked and 15.3 s on sharp curves). Character-based speech recognition took more than twice as long (39.1 s) and times increased slightly when driving (37.1 s while parked and 41.0 s on sharp curves), $t(23) = 2.9$, $p < .01$, suggesting that the speed of speech input was relatively unaffected by the workload levels explored. Older participants performed word-based speech recognition 4 s slower than

TABLE 1: Summary of Activities and Their Sequence for a Typical Participant

	Driving Workload (Curvature)	No. of Addresses
Pretest forms		
Driving		
Practice and baseline	0°, 3°, 9°	
Keyboard entry		
Practice and Baseline 1	parked	12
Practice while driving	0°	6
3 tests while driving	0°, 3°, 9° balanced	18
Baseline 2	parked	6
Break	—	—
Character-based speech recognition		
Practice and baseline	parked	12
2 tests while driving	0°, 9° balanced	12
Word-based speech recognition		
Baseline	parked	6
1 test while driving	9°	6
Driving baseline	0°, 3°, 9°	
Posttest debriefing		

Note. Order of address entry method was reversed for half of the participants. Order of road curvature was counterbalanced based on a Latin square design. Degree of curvature notation: 0° = straight, 3° = moderate curve, and 9° = sharp curve.

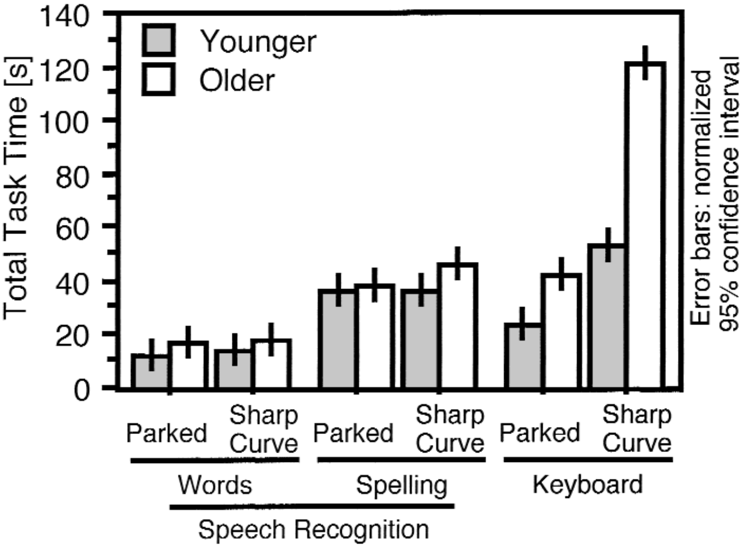

Figure 2. Total task time by address entry method. Error bars represent normalized 95% confidence intervals (CIs) for repeated measures analysis (Masson & Loftus, 2003). As a rule of thumb, plotted means having CIs that overlap by no more than about half the distance of one side of a CI are significantly different.

did younger participants both when parked and on sharp curves, $t(22) = 2.9$, $p < .01$. Although the times for character-based speech recognition were not different between age groups when parked, on sharp curves older participants were 9 s slower than were younger participants, $t(22) = 3.1$, $p < .01$.

While participants were parked, using the touch-screen keyboard (32.5 s) was faster than using character-based speech recognition (37.1 s). On sharp curves, however, keyboard entry took substantially longer than either of the speech recognition modes (83.9 s), as indicated by the interaction between curvature and entry method, $F(2, 40) = 37.6$, $p < .0001$. Overall, total task time on sharp curves was 2.6 times longer than when parked (ratio of 2.9 for older and 2.3 for younger participants). Total task time for typing increased with road curvature significantly, from 32 s while parked to 63 s on straight sections to 84 s on sharp curves, $F(3, 45) = 20.8$, $p < .0001$.

Error Analysis

The probability of errors was experimentally fixed at 8% per word (one error per three addresses) in the speech recognition methods and was measured to be 4.7% per keystroke (about one error per address) when typing on the touch-screen keyboard. The most common typing errors were inadvertent key repetitions (2.6%), pressing adjacent keys (1.3%), or pressing any other keys (0.8%). On average, participants detected an error within 3.5 s of making it (range: 1–16 s).

A detailed analysis was performed on the keystroke data of a sample of four representative participants (randomly selected, one from each age-gender group). Table 2 presents measured parameters, using the equation

$$\begin{gathered}\textit{address entry time} = \\ \sum_{\text{words}} \textit{time per word} + \textit{error probability} \\ \times (\textit{time to detect and} \\ \textit{delete error} + \textit{time to enter correction})\end{gathered} \tag{1}$$

for calculating total task time with consideration of errors. As a demonstration of the detrimental effect of errors on total task times, in word-based speech recognition, detecting and correcting an error increased the total task time by 9 s (+71%).

Glance Timing

Glance analysis for the keyboard-entry task of four addresses (two on a straight road and two on a sharp curve) was examined in detail for 12 participants, 3 from each age-gender group, for whom good-quality video recordings were

TABLE 2: Timing of Address Entry With and Without Errors

	Time per Word (s)	Entry Time With No Error (s)	Error Recovery (s)		Entry Time With 1 Error (s)	Cost of Error (s)
			Detect and Delete	Enter Correction		
Word-based (all)	3.3	13.1	6.0	3.3	22.4	9.3 (+71%)
Character-based						
Parked	8.5	33.9	4.8	9.9	48.6	14.7 (+43%)
Sharp curve	9.5	36.9	7.0	10.0	53.9	17.0 (+46%)
Touch screen						
Parked	7.0	28.0	3.7	3.9	35.6	7.6 (+23%)
Sharp curve	14.7	58.8	4.4	3.0	66.2	7.4 (+11%)

Note. Overall address entry time as a function of error probability can be calculated using Equation 1. Time per word is adjusted for five characters per word. System recognition time is included in entry times.

available. Glances began when the participant's eyes left the road and ended when the participant's eyes left the display to look back at the road (as defined by the Society of Automotive Engineers, 1999). All video analysis was performed by one person.

Table 3 displays data for the three glance measures that were examined. Total glance duration, the overall duration of glances for a single address-entry task, remained relatively unchanged from straight roads to sharp curves. Mean glance durations decreased significantly, from 1.4 to 1.1 s, $F(1, 9) = 53.2$, $p < .0001$, and the mean time between glances increased significantly with road curvature, from 1.1 to 1.4 s, $F(1, 9) = 22.6$, $p < .001$. Consequently, participants looked at the display more of the time on straight sections (60%) than on sharp curves (46%), $F(1, 9) = 27.2$, $p < .001$, and the total number of glances per address on sharp curves increased from 22.0 to 34.2, $F(1, 9) = 17.0$, $p < .01$.

All participants made shorter glances at the display on sharp curves than on straight sections (Figure 3), and most participants (11 of 12) made longer glances at the roadway on sharp curves. Both trends were expected because of the higher visual demand associated with driving on sharp curves.

Measures of Vehicle Control

Measures of vehicle control on the first 10 s (two consecutive 5-s segments) of two addresses on a sharp curve were analyzed for each of the three entry methods. Similar duration and curvature were selected from the no-task condition to allow for an unbiased comparison. A multiple analysis of variance revealed a significant effect of entry method across all dependent measures tested, $F(18, 162) = 4.34$, $p < .0001$, but no age or gender effects.

Lateral control. Degradation of lateral control diminishes safety. Analysis of three measures of lateral control (*SD* of steering wheel angle, *SD* of lateral position, and number of lane departures) showed that when participants entered an address using a keyboard, vehicle control was worse than in the other conditions, but there was no difference in performance between the baseline and speech recognition conditions.

TABLE 3: Mean Values of Glance Measures

	Road Curvature		Age	
	Straight	Sharp	Younger	Older
Total task 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)	1.1	1.4	0.9	1.6
Number of glances	21.9	33.1	20.6	34.5

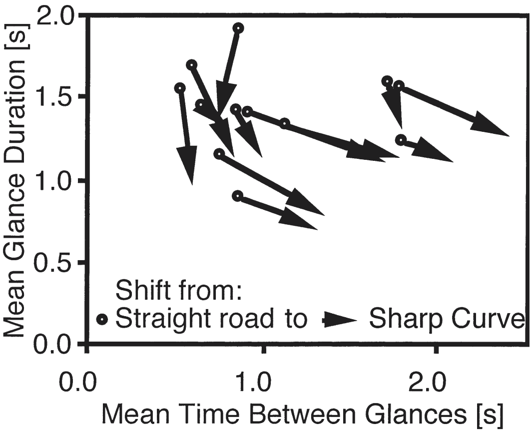

Figure 3. The effect of workload on glance allocation. Each arrow represents 1 participant's performance shift from straight road to sharp curve.

The standard deviation of lateral position when using the keyboard (0.21 m) was 60% higher than that in the other conditions (0.13 m), *F*(3, 57) = 13.4, $p < .0001$ (Figure 4). The standard deviation of steering wheel angle followed the same pattern, increasing from 0.07° when no task was performed and 0.08° on speech recognition (both methods) to 0.12° when using the keyboard, *F*(3, 57) = 15.6, $p < .0001$.

During touch-screen typing, one or more lane departures occurred while participants entered 20.6% of addresses (57/276). The probability of departure was 4.2% (11/263) in character-based speech recognition and 8.3% (12/144) in word-based speech recognition, but they were not statistically different from each other, $\chi^2(1) = 3.0$. In comparison, only 1.5% of the no-task road segments (2/131) had lane departures. The proportion of lane departures during touch-screen typing was significantly more than that in the other three conditions, $\chi^2(1) > 10.4$, $p < .01$, and that for word-based speech recognition was more than that of the baseline, $\chi^2(1) = 6.57$, $p < .05$, but that for spelling-based speech recognition was no different from that for the baseline or for character-based recognition.

Longitudinal control. Analysis of three measures of longitudinal control (accelerator position, forward velocity, and headway) showed a mixed pattern. When participants entered an address, regardless of the entry method, the mean forward velocity was lower than that for baseline driving. The following distance was shortest during word-based speech recognition and longest when using the keyboard. The change in following distance may be a partial compensation mechanism for task overload.

The standard deviation of accelerator position was high when no task was performed (6%), moderate during character-based speech recognition (5%), and low in the other two modes (4%), *F*(3, 57) = 4.8, $p < .01$. The mean forward velocity while participants entered addresses in any of the methods (70.2 kph) was somewhat lower than that while driving only (73.4 kph), *F*(3,

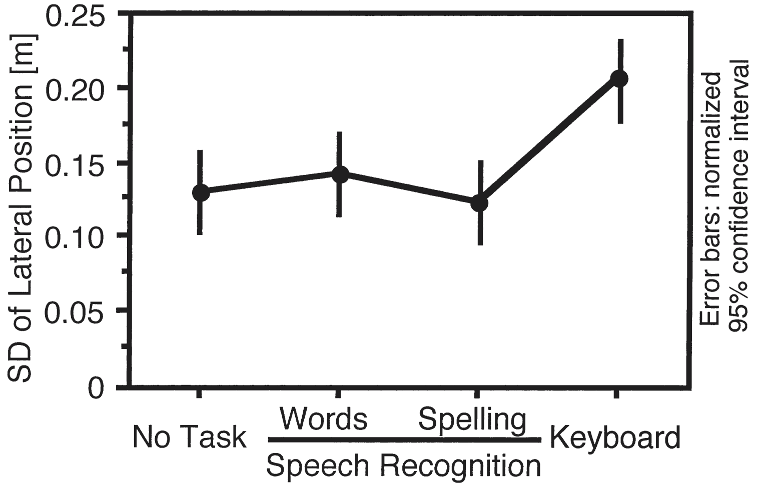

Figure 4. The effect of address entry method on standard deviation of lateral position.

57) = 4.2, *p* < .01 (Figure 5). The standard deviation of forward velocity did not vary significantly between the methods (1.5 kph).

Following distance was shortest during word-based speech recognition (88 m) and longest during keyboard use (167 m), *F*(3, 57) = 2.75, *p* < .05. The standard deviation of the following distance with no task and word-based speech recognition (4.9 m and 4.6 m, respectively) was lower than that for character-based speech recognition and during keyboard use (6.5 and 6.1 m, respectively), *F*(3, 57) = 2.9, *p* < .05.

Difficulty and Safety Ratings

After each experimental block, participants rated task difficulty using a modified Cooper-Harper scale (Wierwille & Casali, 1983), in which 1 = *very easy* and 10 = *impossible.* The effect of entry method, *F*(2, 40) = 39.7, *p* < .0001, the effect of driving, *F*(1, 20) = 63.0, *p* < .0001, and the interaction between them, *F*(2, 40) = 24.6, *p* < .0001, were all significant (Figure 6). Typing the addresses on a keyboard in any of the driving workload levels was rated as significantly more difficult than while parked, *F*(3, 42) = 47.6, *p* < .0001. Difficulty on the sharp curve (7.3) was higher than that on the straight section (5.8), *t*(36) = 2.4, *p* < .05, but not significantly different from that on the moderate curve (6.0). Subjective ratings of spelling-based speech recognition followed the same pattern, significantly increasing with driving workload, *F*(2, 34) = 12.8, *p* < .0001. The mean rating for the parked condition (1.7) was significantly lower than that for the straight (2.8) and the sharp curve (3.4) driving conditions, *t*(40) = 3.8, *p* < .01, but ratings for the two driving conditions were not different from each other. Finally, difficulty ratings of word-based speech recognition were not affected by driving workload, *F*(1, 16) = 0.2.

At the end of the experiment, participants were asked to rate the overall safety of the address entry methods on a 10-point scale (1 = *extremely safe* and 10 = *extremely unsafe*). Word-based speech recognition was rated as relatively safe (4.1), followed by character-based speech recognition (5.3). Keyboard entry was rated as extremely unsafe (9.2), *F*(2, 40) = 47.2, *p* < .0001.

SUMMARY

Total Task Time

As expected, word-based speech recognition was the fastest entry method, followed by character-based speech recognition and keyboard entry. It was not expected, however, that keyboard entry would be faster than character-based speech recognition while participants were parked. The effect of moderate levels of driving workload on the total time for address entry was not significant for word-based speech recognition and was limited in magnitude for

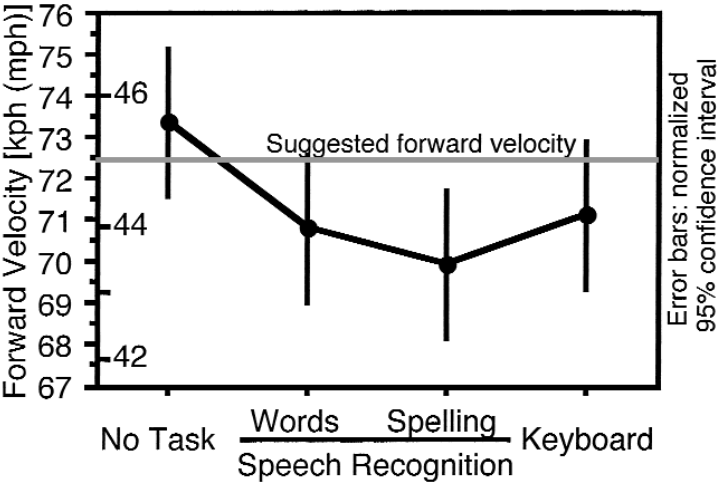

Figure 5. The effect of address entry method on mean forward velocity (in kph on left of axis and in mph on right of axis).

character-based speech recognition, but it was extremely significant when using the keyboard, especially for older participants.

Timing of Glances

Total task time increased with road curvature because of longer glances at the road between glances at the in-vehicle display. As glances at the display became shorter, more glances were made, although the total time looking away from the road increased only slightly.

Degradation of Vehicle Control

Lateral vehicle control when typing an address was significantly worse than that in the baseline no-task condition. As a result, the proportion of trials with at least one lane departure was 20.6% during touch-screen entry as compared with 1.5% in the baseline condition. The proportion of trials with a lane departure during use of one of the speech recognition methods was not as high as that during touch-screen entry but was nearly significantly higher than that for the baseline, suggesting that some interference existed, probably arising from the visual nature of the feedback from the display.

The mean speed dropped by 1.8 kph when participants entered an address using any of the address entry methods. The mean following distance was shortest in word-based speech recognition (88 m) and longest when typing (167 m), suggesting a compensation mechanism to allow for a longer response attributable to distraction from the in-vehicle task.

Perceived Difficulty and Safety

Participants rated the three address entry methods while parked as similarly easy (2.2 on a 10-point scale). For driving tasks, the participants rated the keyboard entry method as most difficult (5.4), followed by character-based speech recognition (3.0) and word-based speech recognition (2.3).

DISCUSSION

Validity of the Driving Task

The driving experience was made as realistic as possible within the limitations of a fixed-based 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. By creating roads of specific, constant curvature in the driving simulator and by controlling the relative position of traffic (up to four cars), we controlled the visual demand of driving, which could not have been done as accurately if the experiment had been performed on the road. Because driving workload levels were controlled, however, the experiment became more sterile than it would have been on the road and the workload levels were limited in range. We believe the benefits of controlled simulation far outweighed this

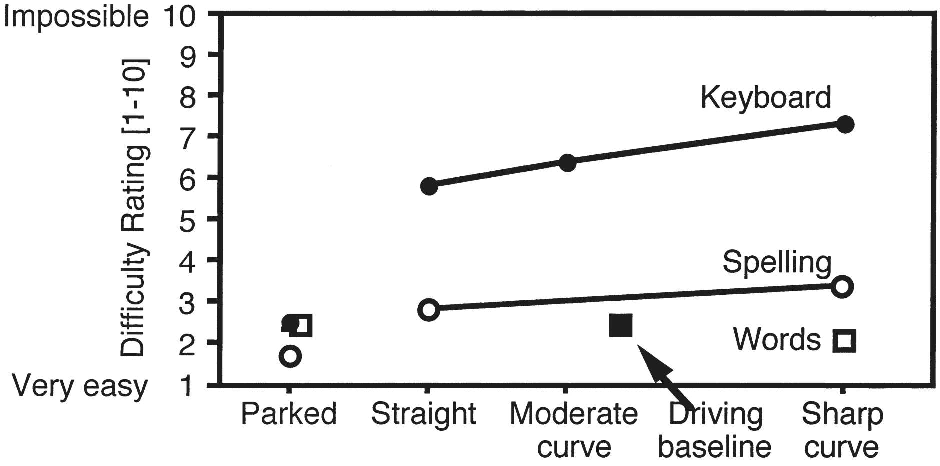

Figure 6. The effect of driving workload on difficulty ratings. For driving baseline, a road consisting of several curvatures was rated. It is therefore displayed at the mean value of the curvatures tested.

limitation for the questions posed. Nevertheless, more emphasis should be given to relative results than to absolute results.

Realism of the Address Entry task

Results and conclusions pertaining to the speech recognition address entry methods are most applicable to systems that use visual, rather than auditory, feedback. Although the address entry methods relied on different modalities (visual and motor for the touch screen and speech for speech recognition), all address entry methods in this experiment used visual feedback. 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 visual feedback.

The simulated touch-screen keyboard resembled some current in-vehicle systems, although it could be improved if it were enlarged and, perhaps, located more conveniently. Another improvement commonly used in contemporary systems is an intelligent speller, which dynamically eliminates characters from the keyboard based on what was previously typed and on what constitutes valid entries.

Finally, participants read the addresses in this experiment from a memo display, not from memory, a piece of paper, or an electronic organizer, as they would in the real world. The question of how drivers retrieve an address they are about to enter into a navigation system was not addressed.

Implications and Applications

This paper reports advances in experimental methods as well as suggestions for improved product design. With regard to methods, we make two recommendations:

- In light of the large performance differences we observed that were attributable to driving workload and visual demand, future experiments should include driving workload as an experimental factor.
- The Wizard of Oz technique was found to be beneficial for simulating a speech recognition system and is recommended for similar experiments in the future. Close attention to details in constructing a seemingly realistic simulated device is required.

Three recommendations apply to product design:

- Address entry using speech recognition with visual feedback is shorter and safer than with a keyboard and is the recommended approach. Although speech recognition systems certainly reduce risk and its duration, not all such systems with visual feedback should be considered risk free. Degradation of vehicle control was apparent in this simulator experiment and is likely to occur on real roads as well.
- Speech recognition errors and typing errors increased total task time considerably. For word-based speech recognition, the added time attributable to a single error was about 70% (9 s). At the design level, special care should be made to reduce the probability of errors and to minimize the consequences of such errors.
- Address entry using a keyboard is unsafe under some conditions. The probability of lane departure during typing was significantly higher than that during the use of speech recognition. Although several young participants typed addresses quickly and without deterioration of vehicle control, other participants were engaged with the typing task for long durations and their vehicle control deteriorated. Regardless of performance level, most participants rated address entry using a keyboard as very difficult and extremely unsafe.

ACKNOWLEDGMENTS

The authors would like to thank Mitsubishi Motors Corporation for funding this research. Special thanks to John Lee, David Strayer, and two anonymous reviewers for their comments.

REFERENCES

Chiang, D. P., Brooks, A. M., & Weir, D. H. (2001). *An experimental study of destination entry with an example automobile navigation system* (SAE Paper 2001-01-0810), Warrendale, PA: Society of Automotive Engineers.

Coleman, M. F., Loring, B. A., & Wiklund, M. E. (1991). User performance on typing tasks involving reduced-size, touch screen keyboards. In *Vehicle Navigation and Information Systems Conference Proceedings* (VNIS '91; pp. 534–549). New York: Institute of Electrical and Electronics Engineers.

Farber, E., Blanco, M., Foley, J., Curry, R., Greenberg, J., & Serafin, C. (2000). Surrogate measures of visual demand while driving. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 3.274–3.277). Santa Monica, CA: Human Factors and Ergonomics Society.

Green, P., & Shah, R. (2003). *Task time and glance measures of the use of telematics: A tabular summary of the literature*

(Tech. Report UMTRI-2003-33), Ann Arbor, MI: University of Michigan Transportation Research Institute.

Green, P., & Wei-Haas, L. (1985). The rapid development of user interfaces: Experience with the Wizard of Oz method. In *Proceedings of the Human Factors Society 29th Annual Meeting* (pp. 470–474). Santa Monica, CA: Human Factors and Ergonomics Society.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based E-mail on drivers' attention to the roadway. *Human Factors, 43,* 631–640.

Masson, M. E. J., & Loftus, G. R (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57,* 203–220.

Nowakowski, C., Utsui, Y., & Green, P. (2000). *Navigation system destination entry: The effects of driver workload and input devices, and implications for SAE recommended practice* (Tech. Report UMTRI-2000-20). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Richardson, B., & Green, P. (2000). *Trends in North American intelligent transportation systems: A year 2000 appraisal* (Tech. Report 2000-9). Ann Arbor, MI: University of Michigan Transportation Research Institute.

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.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9,* 23–32

Tijerina, L., Parmer, E., & Goodman, M. J. (1998). Driver workload assessment of route guidance system destination entry while driving: A test track study. In *Proceedings of the Fifth World Congress on Intelligent Transport Systems* [CD-ROM]. Washington, DC: ITSA.

Tsimhoni, O., & Green, P. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1586–1590). Santa Monica, CA: Human Factors and Ergonomics Society.

Tsimhoni, O., Smith, D., & Green, P. (2001). *Destination entry while driving: Speech recognition versus a touch-screen keyboard* (Tech. Report UMTRI-2001-24). Ann Arbor, MI: University of Michigan Transportation Research Institute.

University of Michigan Transportation Research Institute. (2001). *Driving simulators.* Retrieved September 16, 2004, from www.umich.edu/~driving/sim.html

Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 129–133). Santa Monica, CA: Human Factors and Ergonomics Society.

Wierwille, W. W., & Tijerina, L. (1998). Modeling the relationship between driver in-vehicle demands and accident occurrence. In A. G. Gale, I. D. Brown, C. M. Haslegrave, & S. P. Taylor (Eds.), *Vision in vehicles VI* (pp. 233–243). Amsterdam: Elsevier Science.

Omer Tsimhoni is a research fellow in the Human Factors Division at UMTRI and an adjunct assistant professor of industrial and operations engineering at the University of Michigan, where he received his Ph.D. in industrial and operations engineering in 2004.

Daniel Smith is a human factors engineer at Toyota Technical Center, Ann Arbor, Michigan. He received his M.S.E. in industrial and operations engineering at the University of Michigan in 2004.

Paul Green is a research professor in the Human Factors Division at UMTRI and an adjunct associate professor of industrial and operations engineering and of mechanical engineering at the University of Michigan, where he received a joint Ph.D. in industrial and operations engineering and in psychology in 1979.

Date received: February 5, 2003
Date accepted: July 26, 2004